Jak być człowiekiem

Hello world.

w epoce maszyn

„Przyszłość nie zdarza się tak po prostu. My ją tworzymy”.
Hannah Fry

HANNAH FRY

HANNAH FRY

Hello world

Jak być człowiekiem w epoce maszyn

PRZEŁOŻYŁ
Sebastian Musielak

Wydawnictwo Literackie

Spis treści

Karta redakcyjna

Dedykacja

Parę słów o tytule

Wstęp

Władza

Dane

Wymiar sprawiedliwości

Medycyna

Samochody

Przestępczość

Sztuka

Zakończenie

Podziękowania

Podziękowania za udostępnienie zdjęć

Przypisy

Tytuł oryginału: HELLO WORLD. HOW TO BE HUMAN IN THE AGE OF THE MACHINE

Opieka redakcyjna: MACIEJ ZARYCH
Konsultacja: dr RAFAŁ MICHALCZAK
Redakcja: PAWEŁ CIEMNIEWSKI
Korekta: Pracownia 12A, HENRYKA SALAWA
Projekt okładki: GEOFFREY DAHL
Adaptacja okładki na podstawie oryginału: MAREK PAWŁOWSKI
Redakcja techniczna: ROBERT GĘBUŚ
Skład i łamanie: Infomarket

Copyright © 2018 by Hannah Fry
All rights reserved including the rights of reproduction in whole or in part in any form
© Copyright for the Polish translation by Sebastian Musielak
© Copyright for this edition by Wydawnictwo Literackie, 2019

Wydanie pierwsze

ISBN 978-83-08-06733-8

Wydawnictwo Literackie Sp. z o.o.
ul. Długa 1, 31-147 Kraków
tel. (+48 12) 619 27 70
fax. (+48 12) 430 00 96
bezpłatna linia telefoniczna: 800 42 10 40
e-mail: ksiegarnia@wydawnictwoliterackie.pl
Księgarnia internetowa: www.wydawnictwoliterackie.pl

Konwersja: eLitera s.c.

Dla Marie Fry
Dziękuję, że nie chciałaś słuchać moich wymówek.

Parę słów o tytule

Kiedy miałam siedem lat, tata przyniósł do domu ZX Spectrum, mały, ośmiobitowy komputer. Ani ja, ani moje siostry nie miałyśmy wcześniej komputera. Mimo że był używany i pewnie już od dobrych pięciu lat uważano go za staroć, ja i tak byłam przekonana, że ta skrzynka skrywa w sobie coś wyjątkowego. Nasza czarna bestia była z grubsza zbliżona parametrami do Commodore'a 64 (komputera, który w latach dziewięćdziesiątych miały tylko naprawdę bogate dzieciaki), ale w moim odczuciu prezentowała się znacznie atrakcyjniej. Lśniąca obudowa z tworzywa sama pchała się do rąk, a szare gumowe klawisze i ukośny tęczowy pasek z prawej strony budziły natychmiastową sympatię.

Pojawienie się w naszym domu ZX Spectrum było początkiem niezapomnianych chwil: razem ze starszą siostrą spędziłam wtedy całe lato na strychu, pisząc programiki do gry w szubienicę czy rysowania prostych figur. Na bardziej zaawansowane zabawy trzeba było trochę poczekać. Najpierw musiałyśmy opanować podstawy.

Nie pamiętam dokładnie, kiedy napisałam swój pierwszy w życiu program komputerowy, ale chyba wiem, co to było: ten sam prosty program, którego później uczyłam wszystkich swoich studentów na University College of London; ten sam, który można znaleźć na pierwszej stronie niemal każdego podręcznika podstaw informatyki. Jestem tego pewna, bo w gronie osób, które kiedykolwiek uczyły się programować, istnieje pewna tradycja, swoisty rytuał przejścia: pierwszym zadaniem adepta informatyki jest zaprogramowanie komputera tak, żeby wyświetlił na monitorze słynne już dzisiaj słowa:

HELLO WORLD.

Tradycja ta sięga lat siedemdziesiątych ubiegłego wieku, bo to wtedy właśnie Brian Kernighan zawarł ten programik w swoim niezwykle popularnym podręczniku programowania[1][1*]. Książka ta, a wraz z nią to słynne wyrażenie stały się kamieniami milowymi historii komputerów – na arenie dziejów pojawił się właśnie mikroprocesor, obwieszczając koniec epoki komputerów przeszłości – gigantycznych machin dla wybranych, karmionych przez specjalistów kartami i taśmami perforowanymi – i początek nowej ery, w której w domach zaczęły pojawiać się zbliżone do znanych nam dzisiaj komputery osobiste z monitorem, klawiaturą i migającym kursorem. Program „Hello world" powstał na progu czasów, kiedy pogawędka z własnym komputerem dopiero zaczynała być możliwa.

Wiele lat później Brian Kernighan zdradził dziennikarzowi „Forbesa", skąd wziął inspirację do swojego słynnego programu: otóż zobaczył w jakiejś kreskówce pisklę, które tuż po wykluciu się z jaja zapiszczało radośnie: „Hello world!".

Nie do końca wiadomo, kogo to pisklę miałoby odgrywać w naszym scenariuszu: czy nieustraszonego amatora, który triumfalnie obwieszcza swoje przybycie do świata programistów, czy może sam komputer, budzący się z nudnego snu wypełnionego arkuszami kalkulacyjnymi i edytorami tekstów i odkrywający prawdziwy świat, gdzie przyjdzie mu wykonywać polecenia nowych panów? Być może jedno i drugie. W każdym razie „Hello world" jednoczy wszystkich programistów świata, łącząc ich zarazem z każdą maszyną, jaką zaprogramował człowiek.

Lubię to wyrażenie z jeszcze innego powodu: chyba nigdy nie było bardziej aktualne i ważniejsze niż teraz. W czasach, kiedy algorytmy komputerowe coraz bardziej kontrolują i określają naszą przyszłość, „Hello world" stanowi przypomnienie o możliwości dialogu człowieka z maszyną. O chwili, w której naprawdę nie sposób już ustalić, kto wydaje polecenia, a kto je wykonuje. To początek partnerstwa – wspólnej podróży, której nie zdołamy odbyć w pojedynkę.

W epoce maszyn warto o tym pamiętać.

Wstęp

Każdy, kto kiedykolwiek odwiedził park w Jones Beach na wyspie Long Island w Nowym Jorku, na pewno zauważył wiadukty w drodze nad ocean. Konstrukcje te, zaprojektowane głównie z myślą o kierowaniu pojazdów z autostrady i na autostradę, mają pewną niezwykłą cechę: otóż ich łagodne łuki zawieszone są wyjątkowo nisko nad asfaltową jezdnią – bywa, że zaledwie około 2,8 metra.

To nie przypadek. Cofnijmy się do lat dwudziestych XX wieku. Robert Moses, wpływowy nowojorski urbanista, bardzo chciał przekształcić swój nowy park w Jones Beach w oazę spokoju dla białych i zamożnych Amerykanów. Wiedząc, że preferowana przez niego klientela będzie dojeżdżać na plażę prywatnymi autami, a czarnoskóra biedota publicznym transportem, próbował ograniczyć dostęp tej drugiej do parku, przerzucając setki wiaduktów bardzo nisko nad autostradą. Tak nisko, by wyższe o niespełna metr autobusy nie mogły pod nimi przejechać[1].

Rasistowskie wiadukty to nie jedyne przykłady materii nieożywionej, która bez szumu i rozgłosu sprawuje kontrolę nad ludźmi. Historia zna niezliczone przypadki przedmiotów i wynalazków o możliwościach kontroli znacznie wykraczających poza zakres ich rzekomych zastosowań[2]. Niekiedy możliwości te są z makiaweliczną przebiegłością zaplanowane od samego początku, a czasem znów objawiają się jako wynik bezmyślności projektanta: wystarczy wspomnieć o miejskich torach przeszkód dla wózków inwalidzkich. Czasem znów bywają niezamierzonym skutkiem, jak zmechanizowane warsztaty tkackie z XIX wieku, które miały jedynie ułatwić produkcję tkanin o skomplikowanym splocie, stało się jednak inaczej: całkiem możliwe, że przez to, jak wpłynęły na zarobki, bezrobocie i warunki pracy, pomogły ustanowić tyranię gorszą od kapitalizmu epoki wiktoriańskiej.

Współczesne wynalazki nie stanowią wyjątków. Spytajcie choćby mieszkańców Scunthorpe w północnej Anglii, pozbawionych możliwości

założenia konta na portalu AOL, odkąd ten internetowy gigant uruchomił nowy filtr rodzinny, który miał zastrzeżenia do nazwy ich miasta[3][2*]. Albo Nigeryjczyka Chukwuemekę Afigbo, odkrywcę dozownika mydła do rąk, który działał doskonale za każdym razem, kiedy jego biały kolega umieszczał swoją dłoń pod czujnikiem, nie potrafił już jednak rozpoznać ciemniejszego odcienia skóry[4]. Albo Marka Zuckerberga, który pisząc w 2004 roku kod Facebooka w swoim pokoju w harvardzkim akademiku, nie potrafiłby sobie nawet wyobrazić, że jego dzieło zostanie kiedyś oskarżone o współudział w manipulacjach wyborczych na całym świecie[5].

Za każdym z tych wynalazków stoi algorytm. Algorytmy, niewidzialne fragmenty kodu tworzące konstrukcję i mechanikę współczesnej ery maszyn, dały światu wszystko – od subskrypcji kanałów informacyjnych w mediach społecznościowych, przez wyszukiwarki i nawigację satelitarną, aż po systemy rekomendacji utworów muzycznych – i stanowią część naszej nowoczesnej infrastruktury na równi z mostami, budynkami i fabrykami. Zainstalowaliśmy je w szpitalach, salach sądowych i samochodach. Używane są przez policję, supermarkety i studia filmowe. Poznały nasze upodobania i antypatie; mówią nam, co obejrzeć, co przeczytać i z kim się umówić na randkę. Mają przy tym ukryte możliwości, dzięki czemu powoli i subtelnie zmieniają kryteria człowieczeństwa.

W tej książce poznamy rozmaite algorytmy, którym coraz bardziej – choć zapewne nieświadomie – zawierzamy w życiu. Przyjrzymy się dokładnie temu, co nam obiecują, zrewidujemy ich ukryte możliwości i zmierzymy się z pytaniami, z którymi nas konfrontują. Poznamy algorytmy używane przez policję w celu ustalenia potencjalnych podejrzanych, które każą nam wybierać: czy chronić ofiary przestępstw, czy uniwersalne prawo do domniemywania niewinności. Poznamy algorytmy używane przez sędziów do formułowania orzeczeń, które zmuszają do zastanowienia się nad tym, jak ma wyglądać nasz wymiar sprawiedliwości. Znajdziemy tu algorytmy używane przez lekarzy do podważania własnych diagnoz; algorytmy stosowane w samochodach autonomicznych, które żądają od nas, abyśmy uściślili nasze zasady moralne; algorytmy, które wpływają na to, jak wyrażamy emocje; algorytmy, które mogą podkopywać nasze demokracje.

Nie twierdzę, że algorytmy są z natury złe. Jak się przekonasz, czytając tę książkę, mamy wiele powodów, by na naszą przyszłość patrzeć z optymizmem. Żaden przedmiot ani żaden algorytm nie może być zły czy dobry sam w sobie – o tym decyduje sposób jego wykorzystania. System GPS wymyślono dla systemów nuklearnych, a dziś pomaga nam realizować zamówienia na pizzę. Zapętlona muzyka pop była wykorzystywana w charakterze tortury, a bez względu na to, jak piękny może być wieniec z kwiatów, teoretycznie mogłabym cię nim udusić. Chcąc wyrobić sobie zdanie o algorytmie, musimy najpierw zrozumieć relację człowieka i maszyny. Każdy algorytm wiąże się z ludźmi, którzy go opracowali i potem wdrożyli.

Tak naprawdę zatem jest to książka o ludziach. O tym, kim jesteśmy, dokąd zmierzamy, co jest dla nas ważne – i jak to się zmienia dzięki nowoczesnej technice. To książka o naszej relacji z algorytmami, które już istnieją, które pracują z nami ramię w ramię, zwiększając nasze możliwości, poprawiając nasze błędy, rozwiązując nasze problemy – i przysparzając nam nowych.

Stawiajmy sobie pytanie, czy bilans zysków i strat dla danego algorytmu będzie dla nas korzystny; pytajmy o to, w którym momencie powinniśmy bardziej zaufać maszynie niż własnemu osądowi, a kiedy należałoby oprzeć się pokusie i nie cedować władzy na maszynę. Rozkładajmy algorytmy na czynniki pierwsze i rozpoznawajmy ich ograniczenia; zaglądajmy głęboko w siebie i rozpoznawajmy własne ułomności. Ustalajmy, co nam szkodzi, a co nam pomaga, i decydujmy, w jakim świecie chcielibyśmy żyć.

Przyszłość nie zdarza się po prostu. My ją tworzymy.

Władza

Garry Kasparow doskonale wiedział, jak zastraszać rywali. W wieku trzydziestu czterech lat był największym szachistą, jakiego widział świat – tak porażająca renoma musiała działać na każdego. Na domiar złego stosował jeszcze jeden paskudny trik, który zupełnie odbierał przeciwnikowi chęć do walki: kiedy ten łamał sobie głowę nad zapewne najtrudniejszą partią w całym swoim życiu, Rosjanin niedbałym ruchem sięgał po zegarek, który do tej pory leżał obok szachownicy, i zakładał go z powrotem na przegub lewej ręki. Gest mówił sam za siebie – Garry'emu znudziła się już zabawa w kotka i myszkę. Sztuczka z zegarkiem była jasnym sygnałem dla rywala, że czas złożyć broń. Nie każdy chciał się podporządkować, ale wiadomo było, że partia jest już i tak przegrana[1].

Kiedy jednak w maju 1997 roku doszło do słynnego meczu Kasparowa z zaprojektowanym przez firmę IBM komputerem szachowym Deep Blue, przeciwnik Rosjanina był odporny na podobne zagrywki. Wynik tego pojedynku jest dobrze wszystkim znany, znacznie mniej za to wiadomo, w jaki sposób Deep Blue zdołał go sobie zapewnić. Podstawą tego symbolicznego zwycięstwa maszyny nad człowiekiem, które z wielu względów można uważać za początek epoki algorytmów, było coś więcej niż tylko czysta, surowa moc obliczeniowa. Żeby pokonać Kasparowa, Deep Blue musiał zobaczyć w nim nie tylko nadzwyczaj sprawny procesor szachowy mogący wyliczać rewelacyjne posunięcia, lecz także – człowieka.

Inżynierowie IBM-u błysnęli geniuszem i tak zaprojektowali swój Deep Blue, żeby maszyna sprawiała wrażenie znacznie mniej pewnej swoich obliczeń, niż była w rzeczywistości. Podczas słynnego sześciopartiowego meczu z maja 1997 roku komputer po zakończeniu obliczeń wstrzymywał się co pewien czas z oznajmieniem ruchu, niekiedy nawet na kilka długich minut. Z drugiej strony szachownicy opóźnienia te mogły być odbierane jak wewnętrzne zmagania maszyny, która usiłuje ciągle przeliczać nowe warianty. To z kolei zdawało się potwierdzać przypuszczenia Kasparowa:

że zdołał doprowadzić do sytuacji, w której liczba możliwych wariantów stawała się tak oszałamiająca, iż Deep Blue nie był już w stanie się zdecydować na żaden sensowny ruch[2]. W rzeczywistości Deep Blue doskonale wiedział, jaki ruch wykona; po prostu włączał bieg jałowy i czekał, nie przejmując się upływem czasu. Sztuczka była wredna, ale bardzo skuteczna. Już w pierwszej partii meczu Kasparow zaczął mieć problemy z koncentracją, snując domysły o możliwościach swojego przeciwnika[3].

Ostatecznie pierwszą partię wygrał Kasparow, lecz już w drugiej Deep Blue na dobre zaczął czytać mu w myślach. Rosjanin próbował zwabić komputer w pułapkę, podsuwając mu kilka figur w zamian za doskonałą pozycję – przyjęcie ofiary pozwoliłoby mu po kilku posunięciach włączyć do gry hetmana i rozpocząć atak[4]. Wszyscy obserwujący partię eksperci szachowi spodziewali się – podobnie jak sam Kasparow – że maszyna połknie haczyk. Deep Blue jednak zwęszył podstęp. Ku wielkiemu zaskoczeniu arcymistrza przewidział jego zamiary i zablokował mu hetmana, przekreślając jakiekolwiek szanse na zwycięstwo w tej partii człowieka z maszyną[5].

Kasparow był wyraźnie wstrząśnięty. Błąd w ocenie możliwości komputera zupełnie wytrącił go z równowagi. W wywiadzie udzielonym kilka dni po meczu powiedział, że Deep Blue „przez pewien czas grał, jakby był jakimś bogiem”[6]. Kiedy wiele lat później wrócił myślami do tamtej chwili, napisał tak: „Popełniłem błąd, bo założyłem, że ruchy komputera, które mogły wydawać się zaskakujące, były zarazem obiektywnie mocne”[7]. Tak czy owak, był to triumf geniuszu algorytmu. Deep Blue zrozumiał ludzki umysł – i zarazem ludzką omylność – po czym przypuścił szturm i zwyciężył.

Zdruzgotany Kasparow poddał drugą partię, rezygnując z walki o remis. Już nie odzyskał dawnej pewności siebie. Trzecia, czwarta i piąta partia zakończyły się remisem. Kiedy Kasparow siadał do szóstej partii, był załamany. Mecz zakończył się wynikiem: Deep Blue 3½ – Kasparow 2½.

Dziwna to była przegrana. Rosjanin miał wszelkie szanse, by wyjść obronną ręką z trudnych pozycji na szachownicy, najpierw jednak zbyt nisko oszacował możliwości algorytmu, a potem pozwolił mu się

zastraszyć. „Byłem pod wielkim wrażeniem gry Deep Blue – napisał w 2017 roku. – Za bardzo zacząłem przejmować się tym, do czego jeszcze może być zdolny, i straciłem z oczu fakt, że mam problemy nie dlatego, że on gra tak dobrze, tylko dlatego, że ja gram tak źle”[8].

Jak jeszcze nieraz przekonamy się w tej książce, oczekiwania są ważne. Historia Deep Blue i wielkiego arcymistrza pokazuje, że potęga algorytmu nie jest prostą funkcją tego, co zostało zapisane w liniach jego kodu. Zrozumienie naszych ułomności i słabości oraz ograniczeń maszyny to klucz do utrzymania kontroli.

Jeżeli jednak nie potrafił tego pojąć człowiek pokroju Kasparowa, to jaką szansę mają zwykli zjadacze chleba? Na kolejnych stronach przekonamy się, że algorytmy zakradły się do niemal każdej sfery życia współczesnego człowieka: od służby zdrowia i organów ścigania po transport i politykę. Tymczasem my z jednej strony wypieramy ich istnienie, a z drugiej – żyjemy w bezustannym lęku przed ich nadludzkimi możliwościami. W ostatecznym rozrachunku sami za dobrze nie wiemy, jak wielką władzę oddajemy algorytmom – i czy aby sprawy nie zdążyły już zajść zdecydowanie za daleko.

Mała powtórka

Zanim porozmawiamy o tych problemach, warto może poświęcić chwilę słowu „algorytm” i zastanowić się nad tym, co rzeczywiście ono znaczy. To jedno z tych określeń, które mimo częstego używania niewiele nam mówi – bo i jest niejasne. Oficjalna definicja algorytmu brzmi następująco[9]:

> *algorytm* (rzecz.): krokowa procedura rozwiązania problemu lub osiągnięcia określonego celu, szczególnie przez komputer.

I tyle. Algorytm to po prostu ciąg instrukcji logicznych, które mówią od początku do końca, jak wykonać określone zadanie. W tej obszernej definicji algorytmu zmieściłby się także przepis na placek. Albo wskazówki, jakich udzielisz zagubionemu turyście. Instrukcje z IKE-i, filmiki z YouTube’a pokazujące sposób rozwiązania jakiegoś problemu,

a nawet poradniki – teoretycznie każdy samoistny zestaw instrukcji, jak osiągnąć ściśle określony cel, może być uznany za algorytm.

Tylko że słowa „algorytm" nie używa się w takim ogólnym znaczeniu. Zazwyczaj algorytmy określają coś bardziej konkretnego; owszem, rzecz ciągle sprowadza się do listy kolejno wykonywanych instrukcji, lecz algorytmy, o których będziemy mówić w tej książce, to niemal zawsze obiekty matematyczne. Korzystając z równań, arytmetyki, algebry, analizy, logiki i rachunku prawdopodobieństwa, przetwarzają ciąg operacji matematycznych na kod komputerowy. Karmimy je danymi ze świata rzeczywistego, określamy zadanie i każemy je wykonać. To właśnie algorytmy czynią z informatyki prawdziwą naukę i to one stoją za wieloma najbardziej spektakularnymi osiągnięciami nowoczesnych maszyn.

Liczba istniejących dziś algorytmów jest trudna do oszacowania. Każdy z nich ma własne cele, własne dziwactwa, mocne strony i słabe punkty – brak jest metody, która pozwoliłaby na ich jednoznaczne pogrupowanie. Może zatem warto pokusić się o uproszczenie i podzielić zadania wykonywane przez algorytmy w świecie rzeczywistym na cztery główne grupy[10]:

Priorytetyzacja: tworzenie uporządkowanej listy

Google Search przewiduje, którą stronę chcemy zobaczyć, i tworzy ranking wyświetleń. Netflix proponuje film, który mógłby cię zainteresować. Twój TomTom wybiera dla ciebie najszybszą trasę. Google, Netflix czy TomTom do uporządkowania ogromnej liczby możliwości wyboru wykorzystują matematykę. Również Deep Blue opierał się na algorytmie priorytetyzacji, analizując na każdym kroku wszystkie dopuszczalne ruchy na szachownicy i oceniając, który daje największą szansę na zwycięstwo.

Klasyfikacja: wybór kategorii

Nie miałam jeszcze trzydziestki, kiedy nagle zaczęłam być bombardowana na Facebooku reklamami pierścionków z diamentem. A kiedy w końcu wyszłam za mąż, łaziły za mną po całym internecie reklamy testów ciążowych. Za te drobne uciążliwości mogłam podziękować algorytmom klasyfikującym – ulubieńcom branży marketingowej, niestrudzonym, acz dyskretnym, które uznają cię za osobę zainteresowaną pierścionkami czy testami ciążowymi na podstawie twojej

charakterystyki. (I mogą nawet mieć rację, lecz mimo wszystko jest to denerwujące, kiedy na zebraniu w pracy bez przerwy wyskakują ci na monitorze laptopa reklamy zestawów wspomagających płodność). Istnieją algorytmy, które potrafią automatycznie klasyfikować i usuwać niestosowne treści na YouTubie, algorytmy, które opiszą za ciebie zdjęcia z wakacji, oraz algorytmy, które zeskanują twój odręcznie napisany tekst i rozbiją go na pojedyncze litery alfabetu.

Asocjacja: wyszukiwanie linków

Asocjacja to odnajdywanie i oznaczanie relacji między przedmiotami. Sercem algorytmów dopasowania partnerów, takich jak OKCupid, jest właśnie asocjacja – algorytmy te szukają powiązań między profilami swoich członków i na podstawie wyników sugerują pary. Na podobnej zasadzie działa algorytm rekomendacji firmy Amazon, który kojarzy twoje zainteresowania z zainteresowaniami swoich dotychczasowych klientów. To właśnie on zaintrygował pewnego użytkownika serwisu Reddit, który wkrótce po zakupie na Amazonie kija do baseballa dostał tam osobliwą propozycję: „Może zainteresuje cię ta kominiarka?”[11].

Filtrowanie: wyróżnianie tego, co ważne

Algorytmy czasem muszą usunąć pewne informacje, aby skupić się na tym, co jest dla nich ważne – muszą wyodrębnić sygnał z szumu. Niekiedy robią to dosłownie: algorytmy rozpoznawania mowy, wykorzystywane w takich aplikacjach, jak Siri, Alexa i Cortana, najpierw muszą wyizolować twój głos z tła dźwiękowego, żeby móc potem odkodować, co mówisz. A czasem robią tak w przenośni: Facebook i Twitter filtrują treści, które mają jakiś związek z twoimi zadeklarowanymi zainteresowaniami, żeby dostarczyć ci informacje w pakiecie skrojonym na twoją miarę.

Ogromna większość algorytmów jest budowana tak, by łączyć w sobie kilka wymienionych wyżej funkcji. Weźmy na przykład usługę UberPOOL, która dobiera potencjalnych współpasażerów spośród grupy osób udających się w tym samym kierunku. Znając twoją lokalizację i miejsce docelowe, algorytm musi przefiltrować możliwe trasy, wyszukać miejsca styczności z trasami innych użytkowników jadących w tę samą stronę, a w końcu stworzyć z nich grupę, która zostanie przypisana do ciebie – przydzielając zarazem najwyższy priorytet trasom z najmniejszą liczbą zakrętów, aby przejazd był możliwie najefektywniejszy[12].

Tyle zatem potrafią algorytmy. W jaki sposób udaje się im to wszystko zrobić? No cóż – choć możliwości są praktycznie nieograniczone, można je jakoś pogrupować. W zasadzie sposób podejścia algorytmów do treści podpada pod dwa kluczowe wzorce i z obydwoma będziemy spotykać się w tej książce.

Algorytmy regułowe

Algorytmy tego typu pracują na regułach. Ich instrukcje zostały napisane przez człowieka i dlatego są jasne i klarowne. Można powiedzieć, że algorytmy te stosują się do logiki przepisu na placek. Krok pierwszy: zrób to i to. Krok drugi: jeśli X, to Y. Nie oznacza to wcale, że algorytmy tego rodzaju są proste – wzorzec ten jest na tyle pojemny, że można na nim oprzeć potężne programy.

Algorytmy uczenia maszynowego

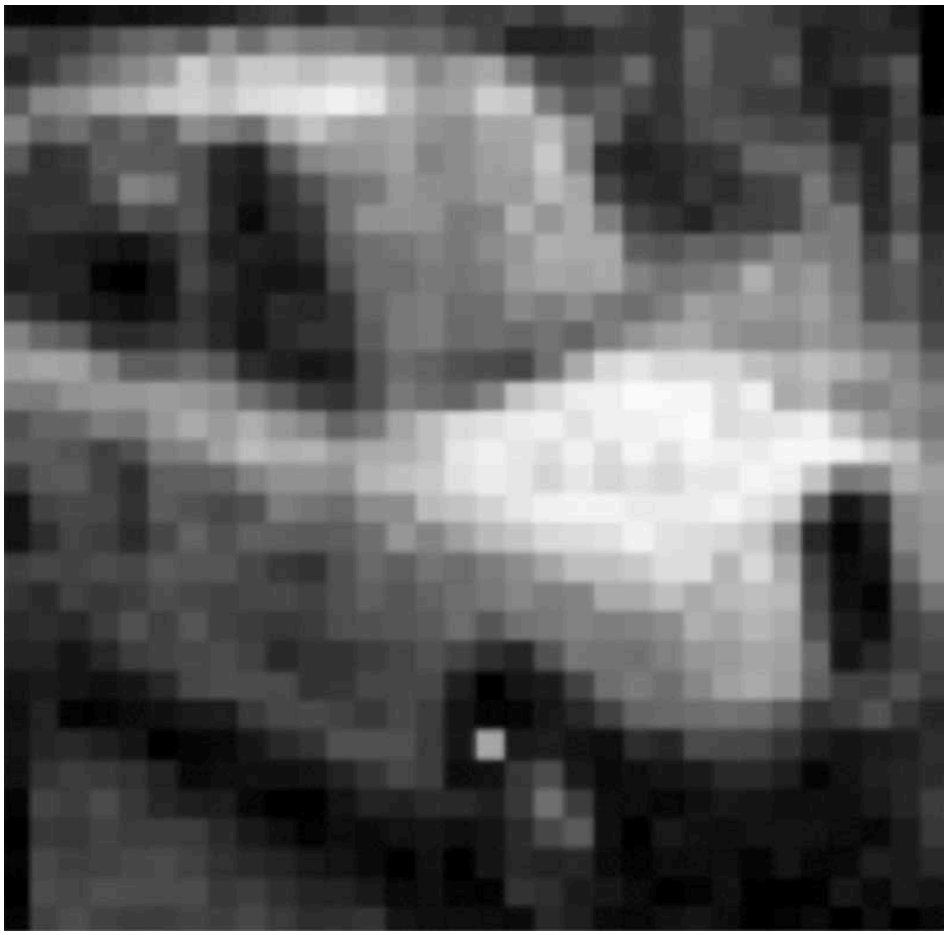
Inspiracją dla tej kategorii algorytmów jest sposób uczenia się istot żywych. Można sobie wyobrazić, że przebiega to analogicznie do procesu uczenia psa, żeby przybijał ci piątkę. Nie musisz tworzyć precyzyjnej listy poleceń i przekazywać ich swojej suczce. Wystarczy, że będziesz dokładnie wiedzieć, co chcesz osiągnąć, i wprowadzisz taki czy inny sposób nagradzania swojej ulubienicy za każdym razem, kiedy zrobi to, o co ją prosisz. Mamy tu zatem do czynienia z prostym mechanizmem wzmacniania właściwych zachowań i ignorowania niewłaściwych w połączeniu z zapewnieniem zwierzęciu czasu niezbędnego do samodzielnego wypracowania własnego trybu działania. Komputerowy ekwiwalent tego procesu nazywany jest a l g o r y t m e m u c z e n i a m a s z y n o w e g o, który zawiera się w szerszym pojęciu s z t u c z n e j i n t e l i g e n c j i (SI). Podajesz maszynie dane, cel oraz informacje zwrotne, kiedy idzie właściwym tropem – i trochę czasu na samodzielne wypracowanie najlepszej metody osiągnięcia wyznaczonego celu.

Oba typy mają swoje plusy i minusy. Algorytmy regułowe wyposażone są w instrukcje napisane przez człowieka i łatwo je zrozumieć. Teoretycznie każdy mógłby zajrzeć do ich kodów i rozgryźć logikę, którą się kierują[13]. Ale ta mocna strona owych algorytmów jest zarazem ich piętą achillesową: algorytmy regułowe będą rozwiązywać jedynie te problemy, dla których człowiek potrafi napisać szczegółową instrukcję.

Algorytmy uczenia maszynowego natomiast udowodniły ostatnio swoje wybitne możliwości rozwiązywania zadań, przy których nie sprawdzają się

napisane przez człowieka listy instrukcji. Algorytmy takie umieją rozpoznawać przedmioty na zdjęciach, rozumieją wymawiane przez nas słowa i tłumaczą je z jednego języka na inny – robią to wszystko, z czym algorytmy regułowe zawsze miały kłopoty. Sęk w tym, że kiedy pozwolisz maszynie samodzielnie poszukać rozwiązania, może się zdarzyć, że ścieżka do niego prowadząca nie będzie miała dla nikogo żadnego sensu. To, co dzieje się wewnątrz takiego algorytmu, stanowi zagadkę nawet dla najbardziej inteligentnego z żyjących programistów.

Weźmy na przykład zadanie polegające na rozpoznawaniu przedmiotów. Znasz pewnie to złudzenie optyczne, kiedy nie potrafisz orzec, czy patrzysz na obraz przedstawiający kielich, czy może dwie twarze (jeśli nie, obrazek znajdziesz w przypisach na końcu książki)[14]. Zespół japońskich badaczy zademonstrował niedawno, jak odmienny od naszego może być algorytmiczny sposób patrzenia na świat. Okazało się, że wystarczyło zmienić tylko jeden piksel na przednim kole samochodu na poniższym zdjęciu, żeby algorytm uczenia maszynowego zmienił zdanie i uznał, iż zdjęcie nie przedstawia samochodu, jak sądził pierwotnie, lecz pokazuje psa[15].



Są tacy, co uważają, że algorytm pracujący bez precyzyjnych instrukcji to przepis na katastrofę. Jak możemy kontrolować coś, czego nie rozumiemy? A jeśli możliwości rozumnych, superinteligentnych maszyn przekroczą możliwości ich twórców? Skąd możemy mieć pewność, że SI, której nie rozumiemy i nie potrafimy kontrolować, już przeciw nam nie spiskuje?

To bardzo interesujące hipotetyczne pytania i bynajmniej nie brakuje literatury poświęconej apokalipsie, jaką nam zgotuje sztuczna inteligencja. Jeżeli właśnie na to po cichu liczysz, to muszę cię rozczarować – ta książka nie będzie o tym. I choć ostatnio SI rozwija się w ekspresowym tempie, nadal jednak jest zaledwie „inteligentna" w najwęższym znaczeniu tego słowa. Obecnie przeżywamy rewolucję bardziej w dziedzinie statystyki obliczeniowej niż inteligencji. Wiem, że to już nie brzmi tak podniecająco (no, chyba że n a p r a w d ę kochasz statystykę), jest to jednak znacznie trafniejszy opis dzisiejszego stanu rzeczy.

Tymczasem więc strach przed złą SI przypomina trochę strach przed przeludnieniem na Marsie[3*]. Być może dożyjemy dnia, w którym inteligencja komputerowa przewyższy inteligencję człowieka, na razie jednak mamy pewność, że nie nastąpi to w najbliższej przyszłości. Szczerze mówiąc, nadal jesteśmy bardzo daleko od stworzenia czegokolwiek o inteligencji jeża. Jak dotąd nikt nie wyszedł poza poziom inteligencji nicienia[4*].

Poza tym zamieszanie wokół sztucznej inteligencji odciąga tylko naszą uwagę od bardziej palących problemów i – przynajmniej dla mnie – znacznie bardziej interesujących historii. Zapomnijmy więc na chwilę o wszechmocnych, inteligentnych komputerach i wróćmy myślami z bardzo odległej przyszłości do naszego tu i teraz, ponieważ już istnieją algorytmy, którym dano prawo podejmowania autonomicznych decyzji – w kwestii długości kary pozbawienia wolności, leczenia pacjentów chorych na raka czy działań, które należy podjąć w chwili zderzenia samochodów. Algorytmy te już dziś, na każdym kroku, decydują za nas o naszym życiu.

Dajemy im szerokie prerogatywy, ale trzeba sobie zadać pytanie: czy algorytmy zasługują na nasze zaufanie?

Ślepa wiara

Niedziela 22 marca 2009 roku nie była najlepszym dniem w życiu Roberta Jonesa. Wracał właśnie od przyjaciół przez piękne miasteczko Todmorden w West Yorkshire, kiedy na desce rozdzielczej jego bmw zapaliła się kontrolka rezerwy. W baku zostało mu paliwa na dziesięć kilometrów z kawałkiem, co nie dawało zbyt dużego pola manewru. Na szczęście jego GPS znalazł dobry skrót – wąską, krętą drogę w górę zbocza doliny.

Robert jechał zgodnie ze wskazówkami urządzenia, droga jednak stawała się coraz węższa i bardziej nachylona. Po paru kilometrach była już zaledwie polną dróżką, na której z trudem wyminęłyby się dwa konie, o samochodach nawet nie wspominając. Ale Robert nie tracił rezonu. Był zawodowym kierowcą, robił osiem tysięcy kilometrów w tygodniu, wiedział więc co nieco o jeździe samochodem. Jak powiedział później: „Nie

miałem żadnych podstaw, żeby nie wierzyć wskazaniom mojego TomToma”[16].

Jeśli chwilę potem ktoś w dolinie przypadkiem spojrzałby w górę, zobaczyłby przednie koła samochodu Roberta wystające znad trzydziestometrowego klifu; od upadku w przepaść uchronił go stojący na skraju urwiska rachityczny drewniany płotek, na którym się zatrzymał.

Potrzeba było traktora i trzech quadów, żeby wyciągnąć zawieszony nad przepaścią samochód. Pod koniec tego samego roku, kiedy Robert stanął przed sądem oskarżony o spowodowanie znacznego zagrożenia w ruchu, przyznał, że nawet nie pomyślał o tym, żeby zignorować wskazania urządzenia. „W nawigacji to była ciągle droga, a nie ścieżka – powiedział w wywiadzie dla prasy. – No więc jechałem według wskazówek. Człowiek się nie spodziewa, że zostanie nagle wyprowadzony na krawędź klifu”[17].

Nie, Robercie. Człowiek się tego nie spodziewa.

Można by rzec, iż to historia z morałem. Mimo że było mu pewnie trochę głupio, że uparcie ignorował to, co widział na własne oczy (zbliżające się urwisko), i zdecydowanie na wyrost przypisywał algorytmowi pewną inteligencję, Robert Jones znalazł się w doborowym towarzystwie. Ostatecznie dwanaście lat wcześniej sam Garry Kasparow dał się nabrać na podobną sztuczkę. W znacznie mniej spektakularnych okolicznościach, choć z równie dramatycznymi konsekwencjami, błąd ten popełnia niemal każdy z nas, zapewne zupełnie nieświadomie.

W roku 2015 naukowcy postanowili ustalić, jak bardzo wyszukiwarki internetowe takie jak Google potrafią zmienić nasz sposób postrzegania świata[18]. Chcieli sprawdzić, czy mamy jakieś zdrowe granice zaufania do wyników wyszukiwań, czy też beztrosko podążamy za nimi aż na krawędź metaforycznego urwiska.

Scenariusz eksperymentu uwzględniał nadchodzące wybory w Indiach. Badacze pod kierunkiem psychologa Roberta Epsteina dobrali 2150 niezdecydowanych wyborców z całego kraju i udostępnili im specjalnie skonstruowaną wyszukiwarkę o nazwie Kadoodle, która miała im pomóc lepiej poznać kandydatów i tym samym zdecydować się, na którego z nich oddać swój głos.

Kadoodle był zmanipulowany. Uczestnicy eksperymentu nie wiedzieli, że zostali podzieleni na grupy i każdej z nich wyszukiwarka pokazywała nieco inną wersję wyników, bo faworyzującą innego kandydata. Kiedy członkowie grupy odwiedzali stronę Kadoodle'a, wszystkie linki na samej górze listy wyników dotyczyły przypisanego do niej polityka; użytkownicy przeglądarki musieliby pominąć wiele linków z pierwszej strony wyszukiwań, żeby znaleźć stronę dającą pozytywny obraz innego kandydata.

Nie będzie niespodzianką, że uczestnicy eksperymentu większość czasu spędzali na przeglądaniu stron internetowych z górnej połówki pierwszej strony wyników – jak w starym kawale: „Gdzie najlepiej ukryć trupa? Na drugiej stronie wyników Google'a!". Prawie wszyscy zignorowali linki, które pojawiały się na dalszych miejscach listy wyników. Niemniej stopień, w jakim faworyzujące określonych kandydatów uporządkowanie linków wpływało na opinie uczestników eksperymentu, zaszokował nawet Epsteina. Kiedy po zaledwie kilku minutach oglądania spreparowanych wyświetleń Kadoodle'a uczestnikom zadano pytanie, na kogo zagłosowaliby w wyborach, prawdopodobieństwo wybrania kandydata faworyzowanego przez wyszukiwarkę wzrosło o zatrważające 12 procent.

W wywiadzie zamieszczonym w czasopiśmie „Science" w 2015 roku[19] Epstein wyjaśnił mechanizm tego zjawiska: „Zakładamy, że wyszukiwarki podejmują słuszne decyzje. Uczestnicy eksperymentu mówią nam: «Tak, widzę, że wyniki nie są obiektywne, i to mi mówi, że (...) wyszukiwarka spełnia swoje zadanie»". Jeśli wziąć pod uwagę, jak wiele informacji czerpiemy dziś z rekomendacji takich właśnie algorytmów wyszukiwarek, jeszcze bardziej przeraża to, jak dalece uczestnicy eksperymentu wierzyli w niezależność swoich opinii: „Ludzie niezdający sobie sprawy z tego, że są manipulowani, skłaniają się ku przekonaniu, że zmieniają poglądy pod wpływem samodzielnych decyzji i przemyśleń", napisał Epstein w opisie swoich badań[20].

Kadoodle nie jest oczywiście jedynym algorytmem, który zgrzeszył subtelnym manipulowaniem sympatiami politycznymi swoich użytkowników. Więcej będzie o tym mowa w rozdziale *Dane*, tymczasem warto pamiętać o tym, co zdaje się sugerować eksperyment Epsteina: mamy skłonność do ufania algorytmom niemal bez zastrzeżeń. Prędzej czy

później uznajemy, że komputer będzie zawsze mądrzejszy od nas[21]. I z upływem czasu przestajemy być nawet tego świadomi.

Wszechobecne dziś algorytmy są dla nas wygodnym autorytetem. Łatwym obiektem do cedowania odpowiedzialności – skrótem, na który się decydujemy bez chwili głębszego namysłu. Bo komu będzie się chciało przy każdym wyszukiwaniu klikać w linki z drugiej strony wyników Google'a, żeby odnieść się krytycznie do każdego z nich? Albo sprawdzać na stronach wszystkich linii lotniczych, czy Skyscanner rzeczywiście pokazał najtańsze przeloty? Mierzyć linijką trasy na mapie, żeby się upewnić, że nawigacja faktycznie wskazuje najkrótszą drogę do wyznaczonego celu? Mnie na pewno by się nie chciało.

Musimy jednak wprowadzić ważne rozróżnienie. Bo zaufanie do zazwyczaj wiarygodnego algorytmu to jedno, a zaufanie do algorytmu bez pewnej wiedzy o jakości jego działania to już zupełnie inna para kaloszy.

Sztuczna inteligencja kontra naturalna głupota

W 2012 roku pewna liczba osób niepełnosprawnych z Idaho otrzymała informację o planowanych ograniczeniach dostępu do programu Medicaid[22]. Mimo że wszyscy zostali zakwalifikowani do otrzymania federalnej pomocy medycznej, władze stanu – bez ostrzeżenia – zdecydowały się na ograniczenie ich świadczeń aż o 30 procent[23], zmuszając pacjentów do samodzielnego szukania środków na pokrycie kosztów leczenia. Nie była to decyzja polityczna; był to efekt wdrożenia przez Departament Zdrowia i Pomocy Socjalnej stanu Idaho nowego „instrumentu budżetowego" – programu komputerowego do obliczania wysokości świadczeń dla beneficjentów programu Medicaid[24].

Problem z „instrumentem budżetowym" był taki, że jego decyzje okazały się zupełnie bezsensowne. Zdaniem postronnych obserwatorów wyliczane przez niego kwoty wyglądały na wzięte z sufitu. Niektórym świadczeniobiorcom przyznano więcej środków niż w poprzednich latach, inni znów stwierdzali, że ich roczny budżet został obcięty o dziesiątki tysięcy dolarów, co groziło eksmisją i koniecznością szukania sobie miejsca w domu pomocy społecznej[25].

Mieszkańcy Idaho nie mogli zrozumieć, dlaczego niektóre ich świadczenia zostały cofnięte, i nie chcąc się pogodzić z taką decyzją, zwrócili się o pomoc do Amerykańskiej Unii Wolności Obywatelskich (ACLU). Ich sprawą zajął się Richard Eppink, dyrektor prawny wydziału Unii w Idaho[26], który napisał o tym w swoim blogu w 2017 roku: „Sądziłem, że sprawa zakończy się wysłaniem prostego zapytania do władz stanowych: «Dobra, powiedzcie nam, dlaczego te kwoty tak bardzo spadły?»[27]. W rzeczywistości potrzeba było aż czterech lat i pozwu zbiorowego podpisanego przez cztery tysiące osób, żeby w końcu ustalić, co się wydarzyło”[28].

Eppink ze swoim zespołem zaczął od pytań o szczegóły działania algorytmu, lecz strona Medicaid odmówiła wyjaśnień na temat metody wyliczania świadczeń, argumentując, że służący do tego program stanowi tajemnicę handlową i dlatego nie może zostać poddany analizie[29]. Na szczęście sędzia prowadzący sprawę nie zgodził się i strona stanowa przekazała do analizy swój instrument budżetowy, któremu dała taką władzę nad obywatelami. Okazało się, że nie był to wcale bardzo skomplikowany, inteligentny program bazujący na dogłębnie przemyślanym modelu matematycznym, lecz zwykły arkusz kalkulacyjny Excela[30].

Obliczenia arkusza opierały się podobno na autentycznych przypadkach, program był jednak obarczony tak poważnymi wadami i błędami, że uznano go niemal w całości za bezużyteczny[31]. Co gorsza, kiedy zespół ACLU przeanalizował formuły arkusza, stwierdził „fundamentalne błędy statystyczne w samej strukturze programu”. Instrument budżetowy na podstawie ogromnej bazy danych świadczeniobiorców generował najzupełniej przypadkowe wyniki. Algorytm – jeżeli można to coś w ogóle nazwać tym słowem – okazał się tak kiepski, że sąd ostatecznie uznał jego użycie za niekonstytucyjne[32].

Biegną tutaj równolegle dwa wątki ludzkiej omylności. Po pierwsze, ktoś musiał stworzyć ten beznadziejny arkusz kalkulacyjny, i po drugie – musieli znaleźć się naiwni, którzy uwierzyli w jego możliwości. „Algorytm” był w istocie spartaczoną robotą człowieka opakowaną w kod

komputerowy. Dlaczego więc władze stanu Idaho tak zaciekle broniły czegoś tak beznadziejnego?

Oto fragment przemyśleń Eppinka: „Wszystko przez to, że nie kwestionujemy wyników przepuszczonych przez komputer. Kiedy maszyna coś generuje (...) po prostu wierzymy na ślepo formule i nie pytamy: Ale zaraz, jak to faktycznie działa?”[33].

Wiem, że branie pod lupę formuł matematycznych, żeby sprawdzić, jak działają, jest bardzo rzadkim hobby (wiem, bo to akurat mój konik), niemniej Eppink podnosi nadzwyczaj ważną kwestię: zwraca uwagę na naszą gotowość do brania algorytmów na wiarę, bez żadnego zainteresowania tym, co się dzieje za kulisami.

Jestem matematyczką, która od lat ma do czynienia z danymi i algorytmami, przekonałam się więc, że jedynym sposobem obiektywnej oceny wiarygodności takich programów jest rozebranie ich na czynniki pierwsze i przeanalizowanie mechanizmu ich działania. Z doświadczenia wiem, że algorytmy bardzo przypominają sztuczki iluzjonisty: na pierwszy rzut oka to prawdziwe czary, ale gdy tylko znasz trik, na którym numer się zasadza, magia pryska od razu jak mydlana bańka. Często za tą fasadą kryje się prostota wręcz porażająca – ale niekiedy także przerażająca – beztroska. W kolejnych rozdziałach, w których będziemy się przyglądać innym algorytmom, postaram się w miarę możliwości uchylić rąbka tajemnicy i wyjaśnić, co się dzieje za kulisami. To powinno wystarczyć, żeby się przekonać, jak się te sztuczki robi, choć może to trochę za mało, by próbować robić je samemu.

Mimo to nawet najbardziej zagorzałym fanom matematyki będą się trafiać algorytmy, którym po prostu trzeba zawierzyć na ślepo. Być może dlatego że – podobnie jak w przypadku wyników wyszukiwań Skyscannera czy Google’a – sprawdzenie poprawności ich działania jest niewykonalne. Albo może dlatego że – jak w przypadku instrumentu budżetowego z Idaho czy innych podobnych narzędzi, które napotkamy na swojej drodze – algorytm okaże się tajemnicą handlową. Czy też dlatego że – jak w niektórych technikach uczenia maszynowego – prześledzenie procesów

logicznych zachodzących w algorytmie po prostu przekracza możliwości ludzkiego umysłu.

Będą przypadki, kiedy trzeba będzie zrzec się kontroli i zdać na niewiadomą – chociaż dobrze wiemy, że algorytmy nie są nieomylne. Będą i takie, kiedy będziemy zmuszeni skonfrontować nasz osąd z osądem maszyny. Będą wreszcie takie sytuacje, gdy – jeżeli już bardziej zaufamy własnemu instynktowi niż obliczeniom komputera – będziemy potrzebować sporo odwagi, żeby pozostać na obranym kursie.

Kiedy zawetować decyzję komputera

Rosyjski oficer Stanisław Pietrow odpowiadał za monitorowanie systemu wczesnego ostrzegania dla radzieckiej strefy powietrznej. W razie gdyby komputer zarejestrował cokolwiek mogącego wskazywać na atak ze strony Amerykanów, Pietrow miał natychmiast zaalarmować swoich przełożonych[34].

Pietrow był na służbie 26 września 1983 roku, kiedy tuż po północy zawyły syreny alarmowe. Tego sygnału bali się wszyscy. Radzieckie satelity wykryły nieprzyjacielski pocisk zmierzający w stronę terytorium ZSRR. Na ówczesnym, mrocznym etapie zimnej wojny uderzenie Amerykanów było jak najbardziej prawdopodobne, Pietrow jednak się zawahał; nie miał pewności, czy może zaufać algorytmowi. Komputery wykryły zaledwie pięć pocisków – irracjonalnie skromną salwę jak na amerykański atak atomowy[35].

Pietrow zamarł w fotelu. Wszystko było teraz w jego rękach: zgłosić alarm i niemal na pewno rozpętać światową wojnę nuklearną albo zignorować protokół i czekać ze świadomością, że z każdą upływającą sekundą ZSRR ma coraz mniejsze szanse na skuteczny kontratak.

Na szczęście dla nas wszystkich Pietrow wybrał to drugie. Nie mógł wiedzieć na pewno, że alarm jest fałszywy, lecz po dwudziestu trzech minutach – musiały być dla niego prawdziwą wiecznością – kiedy stało się jasne, że na rosyjską ziemię nie spadła żadna rakieta z głowicą jądrową, mógł odetchnąć z ulgą, że postąpił słusznie. Algorytm się pomylił.

Gdyby system działał całkowicie autonomicznie, bez nadzoru człowieka, nasza historia bez wątpienia potoczyłaby się inaczej. ZSRR z pewnością wystrzeliłby rakiety i wierząc, że odpowiada na amerykańską agresję, rozpętałby totalną wojnę nuklearną. Jeśli można z tej historii wyciągnąć jakąś naukę, to brzmi ona tak: element ludzki musi stanowić integralną część całego procesu; ustanowienie kontrolera z prawem weta, władnego oceniać sugestie algorytmu przed podjęciem ostatecznej decyzji, to jedyny sensowny sposób na uniknięcie katastrofalnych błędów.

Ostatecznie tylko człowiek będzie odczuwał ciężar odpowiedzialności za własne decyzje. Algorytm obarczony zadaniem zaalarmowania Kremla o zbliżającym się ataku nuklearnym nie pomyślałby nawet przez sekundę o potencjalnych skutkach swojej decyzji. A Pietrow? „Doskonale wiedziałem, że jeśli się pomylę, już nikt nie będzie w stanie naprawić mojego błędu”[36].

Z tą konkluzją jest jednak pewien mały problem – ludziom także nie zawsze można zaufać. Czasami, tak jak Pietrow, słusznie postąpią, ignorując algorytm. Częściej jednak najlepiej jest dla nas, kiedy ignorujemy podszepty intuicji.

Posłużę się innym przykładem ze świata zabezpieczeń. To na szczęście bardzo rzadka historia – człowieka, który zdecydował się nie posłuchać algorytmu, bo takie właśnie było podłoże słynnej tragedii kolejki górskiej Smiler w Alton Towers, największym parku rozrywki w Wielkiej Brytanii[37].

W czerwcu 2015 roku do parku wezwano z powodu awarii kolejki dwóch mechaników, którzy po naprawieniu usterki wysłali pusty skład na próbną jazdę – nie zorientowali się jednak, że wagoniki nie wróciły do punktu startu. Z niewyjaśnionych przyczyn pusty skład stoczył się po pochyłości i zatrzymał na trasie.

W tym czasie, o czym mechanicy nie wiedzieli, rosnące kolejki spowodowały, że obsługa Smilera podstawiła dodatkowy skład. I kiedy dostała ze sterowni zgodę na start, zaczęła wpuszczać rozradowanych pasażerów do wagoników i zapinać im pasy – zupełnie nieświadoma tego, że nieco dalej na torze stoi zapomniany pusty skład wysłany na próbną jazdę przez mechaników.

Na szczęście projektanci toru przewidzieli podobną sytuację i algorytmy bezpieczeństwa zadziałały dokładnie tak, jak zostały zaprogramowane. Aby uniknąć nieuchronnej kolizji, komputer zatrzymał skład z pasażerami przed szczytem pierwszego podjazdu i uruchomił alarm w sterowni. Mechanicy jednak – przekonani, że naprawili usterkę – uznali ów automatyczny alarm za wynik błędu w systemie.

Obejście zabezpieczeń rollercoastera nie jest łatwe: dwie osoby muszą jednocześnie wcisnąć przycisk restartu. Wypełniony do ostatniego miejsca skład ruszył więc w dół toru, by po chwili zderzyć się ze stojącą na drodze przeszkodą. Skutki były tragiczne. Kilkanaście osób odniosło ciężkie obrażenia, a dwóm nastolatkom amputowano nogę.

Oba scenariusze balansowania na granicy życia i śmierci – Alton Towers i alarm Pietrowa – dobitnie ilustrują dylemat o znacznie głębszym znaczeniu. Do kogo – czy może: czego? – powinno należeć ostatnie słowo w tym sporze człowieka i algorytmu o kompetencje?

Przeciąganie liny

Ta dyskusja ma długą historię. W 1954 roku Paul Meehl, profesor psychologii klinicznej z University of Minnesota, stał się zakałą całego pokolenia. Opublikował bowiem książkę pod tytułem *Clinical vs Statistical Prediction* (Diagnoza kliniczna a diagnoza statystyczna), w której zdecydowanie poparł jedną stronę sporu[38].

Systematycznie porównał w niej skuteczność ludzi i algorytmów w szerokim zakresie prognostyki – od przewidywania ocen końcowych studentów po wyniki rozmaitych badań diagnostycznych – i na tej podstawie stwierdził, że algorytmy matematyczne, nawet te najprostsze, są niemal bez wyjątku lepszymi prognostami od ludzi.

Konkluzje Meehla zostały potwierdzone w niezliczonych badaniach, które przeprowadzono w kolejnym półwieczu. Jeżeli rozwiązanie zadania będzie wymagało jakichkolwiek obliczeń – czy będzie to diagnoza lekarska czy prognoza sprzedaży, przewidywanie prób samobójczych czy satysfakcji pracowników, ocena zdolności do służby wojskowej czy symulacja przyszłej średniej na uczelni wyższej – możesz bez obawy postawić

wszystkie pieniądze na algorytm[39]. Maszyna nigdy nie będzie doskonała, ale danie człowiekowi prawa weta wobec obliczeń algorytmu to potęgowanie błędu ludzkiego[5*].

W zasadzie nie powinno nas to dziwić. Nie zostaliśmy przecież stworzeni do liczenia. W supermarketach nie znajdziemy kasjerek wpatrujących się wytrzeszczonymi oczami w nasze zakupy, żeby ocenić, ile powinniśmy zapłacić – mamy za to (niewiarygodnie prosty) algorytm, który oblicza to za nas. I w większości przypadków lepiej rzeczywiście zdać się na maszynę. Przypomina mi się powiedzenie pilotów cywilnych, że najlepsza załoga kokpitowa liczy trzech członków, są to: pilot, komputer i pies. Komputer ma sterować samolotem, pilot ma karmić psa, a pies ma ugryźć człowieka za każdym razem, kiedy ten tylko spróbuje dotknąć komputera.

Nasza relacja z maszynami ma cechy paradoksu. Niby skłaniamy się ku temu, by naiwnie ufać wszystkiemu, czego nie rozumiemy, ale gdy tylko zdamy sobie sprawę z tego, że algorytmy m o g ą się mylić, wyłazi z nas irytujący histeryk, który całkowicie je odrzuca i zaraz chce zastępować mocno niedoskonałym, ludzkim rozumowaniem. Zjawisko to badacze nazywają awersją algorytmową. Człowiek znacznie gorzej toleruje błędy algorytmu niż swoje własne – nawet jeżeli jego pomyłki są o wiele poważniejsze od pomyłek maszyny.

Istnienie awersji algorytmowej wykazano w licznych eksperymentach[40] i niewykluczone, że pewne jej objawy zaobserwowaliście u siebie. Za każdym razem, kiedy mój Citymapper mówi mi, że dotarcie na miejsce będzie trwało dłużej, niż moim zdaniem powinno, czuję, że chyba wiem lepiej – chociaż najczęściej kończy się to spóźnieniem. Kto przynajmniej raz w życiu nie zwymyślał Siri od idiotek – bo nawet nie pomyślał, jak nieprawdopodobnym osiągnięciem technologicznym było stworzenie mówiącego asystenta, którego można zamknąć w dłoni? Ile to razy w pierwszych tygodniach korzystania z Waze'a pakowałam się w końcu w korek, bo zdołałam sobie wmówić, że jeśli machnę ręką na nawigację i pojadę bocznymi ulicami, to dojadę szybciej? (Prawie nigdy nie dojechałam szybciej). Teraz już jestem mądrzejsza i – jak Robert Jones w swoim bmw – bez szemrania wykonuję polecenia mojej aplikacji, nie

zastanawiając się nawet nad tym, którędy każe mi jechać (choć mimo wszystko nie wjechałabym pewnie na skraj urwiska).

Ta nasza ludzka skłonność do patrzenia zerojedynkowego – uznawanie algorytmów albo za wszechmocne, albo za bezużyteczne – stanowi poważny problem w epoce zaawansowanych technologii. Jeśli chcemy czerpać z jej osiągnięć maksimum korzyści, musimy się nauczyć patrzeć na świat bardziej obiektywnie. Musimy wyciągnąć wnioski z porażki Kasparowa i przyznać się do tego, że jesteśmy omylni; musimy nabrać dystansu do głębokich przeczuć i lepiej pojąć nasz stosunek do wszechobecnych algorytmów – i zarazem zdjąć algorytmy z piedestału, wnikliwiej się im przyjrzeć i postawić pytanie, czy rzeczywiście potrafią aż tyle, ile deklarują. Tylko tak się przekonamy, czy zasługują na władzę, którą im przekazaliśmy.

Niestety, w praktyce sprawa nie jest taka prosta. Bardzo często nie będziemy mieli nic do powiedzenia w kwestii władzy i uprawnień otaczających nas algorytmów – nawet w przypadku tych, których działanie dotyka nas bezpośrednio.

Mowa tu zwłaszcza o algorytmach, które obracają najbardziej chodliwym towarem naszych czasów, a mianowicie: danymi. O algorytmach, które za cichym przyzwoleniem decydentów śledzą nas w internecie i gromadzą o nas informacje, ingerują w naszą prywatność i nas profilują – a wszystko po to, żeby subtelnie oddziaływać na nasze zachowania. Ta wybuchowa mieszanka władzy, wpływów i źle ulokowanego zaufania może radykalnie odmienić nasze społeczeństwa.

Dane

Cofnijmy się do roku 2004, kiedy Mark Zuckerberg – wówczas jeszcze student college'u, który właśnie stworzył Facebook – wymienił z kolegą kilka wiadomości na komunikatorze internetowym:

> „ZUCK: Także jakbyś potrzebował info o kimś z Harvardu…
> ZUCK: To wal.
> ZUCK: Mam ponad 4000 maili, zdjęć, adresów…
> [KOLEGA ZUCKA]: Co? Jak ty to zrobiłeś?
> ZUCK: Sami mi przysyłali.
> ZUCK: Nie wiem czemu.
> ZUCK: Ufają mi.
> ZUCK: Debile"[1].

Słowa te wypływały w 2018 roku wielokrotnie na falach skandalu politycznego wokół Facebooka; dziennikarze forsowali tezę o makiawelicznym stosunku firmy do prywatności użytkowników. Moim zdaniem moglibyśmy w naszych interpretacjach przechwałek dziewiętnastolatka wykazać się trochę większą wspaniałomyślnością. Ale uważam też, że Zuckerberg się mylił. Ludzie nie przesyłali mu swoich danych osobowych ot tak, po prostu. Był to dla nich barter – w zamian za swoje dane otrzymali dostęp do algorytmu pozwalającego im na swobodny kontakt z przyjaciółmi i rodziną; dostali własną przestrzeń, w której mogli się dzielić swoim życiem z innymi. Otrzymali własną prywatną sieć w bezkresnej przestrzeni World Wide Web. Nie wiem jak ty, ale ja wtedy, w tamtym czasie, uważałam, że to naprawdę niezły układ. Ta logika ma jednak pewną ukrytą wadę: nie zawsze potrafimy sobie uzmysłowić długoterminowe implikacje podobnego barteru. Rzadko jest dla nas jasne, co nasze dane potrafią, czy też jak wielką mogą mieć wartość dla twórców

takiego czy innego sprytnego algorytmu. Innymi słowy: nawet nie wiemy, jak tanio się sprzedaliśmy.

Dużo, tanio, Tesco

Supermarkety jako jedne z pierwszych zdały sobie jasno sprawę z wartości indywidualnych profili klientów. W sektorze, gdzie firmy bezustannie walczą o uwagę klienta, każda minimalna poprawa jakości może się przełożyć na ogromny zysk. Tym właśnie powodowała się brytyjska sieć Tesco, kiedy w 1993 roku wprowadziła swoją przełomową kampanię pilotażową.

Pod okiem zespołu kierowanego przez małżeństwo Edwinę Dunn i Clive'a Humby'ego w wybranych marketach Tesco zaczęto oferować klientom zupełnie nową kartę klubową – pokazywany przy kasie plastik, rozmiarami i kształtem przypominający kartę kredytową. Zasady gry były proste: za każdą transakcję z wykorzystaniem karty klubowej Tesco klient otrzymywał punkty obniżające wartość przyszłych zakupów, a firma – możliwość rejestrowania zakupów i kojarzenia ich z nazwiskiem klienta[2].

Dane zebrane podczas tej pierwszej kampanii były bardzo skąpe – oprócz imienia, nazwiska i adresu klienta komputer rejestrował tylko wydaną na zakupy sumę oraz dzień zakupu, ignorując przy tym zupełnie zawartość koszyka. A przecież mimo tak skromnego zasiewu Dunn i Humby zebrali fenomenalne plony.

Odkryli bowiem, że ogromny procent całej sprzedaży generuje niewielka garstka lojalnych klientów. Znali ich adresy, mogli więc prześledzić, z jak daleka byli oni gotowi przyjechać do swojego sklepu. Dowiedzieli się, w których rejonach miast Tesco przegrywa z konkurencją, a w których nad nią góruje. Zgromadzone dane pozwoliły stwierdzić, którzy klienci przychodzą do sklepu codziennie, a którzy przyjeżdżają na zakupy wyłącznie w weekendy. Ta wiedza pozwoliła zespołowi Dunn i Humby'ego opracować plan sterowania zachowaniem konsumentów – za pośrednictwem poczty firma rozesłała posiadaczom karty klubowej Tesco specjalne kupony. Klienci wydający najwięcej otrzymali vouchery na kwotę od 3 do 30 funtów, a ci, którzy wydawali nieco mniej, dostali na zachętę mniejsze kwoty – od 1 do 10 funtów. Rezultaty akcji przerosły najśmielsze

oczekiwania. Niemal 70 procent kuponów zostało zrealizowanych, a ich właściciele wracali do kas z pełnymi koszykami: posiadacze karty klubowej zostawili w sklepach łącznie o 4 procent pieniędzy więcej niż klienci, którzy karty nie mieli.

Dwudziestego drugiego listopada 1994 roku Clive Humby zaprezentował wyniki kampanii pilotażowej na zebraniu zarządu spółki Tesco. Pokazał zgromadzone dane, wskaźniki odpowiedzi, dowody satysfakcji konsumentów, doskonałe wyniki sprzedaży. Zarząd wysłuchał go w milczeniu. Jako pierwszy po prezentacji odezwał się prezes zarządu. „Jestem przerażony. Po trzech miesiącach wiecie więcej o moich klientach niż ja po trzydziestu latach”[3].

Karty klubowe wydano wszystkim klientom i w powszechnej opinii to one właśnie pozwoliły Tesco zostawić w tyle swojego głównego rywala, spółkę Sainsbury’s, i zostać największą siecią supermarketów w Wielkiej Brytanii. Z upływem czasu rejestrowano coraz więcej danych, co jeszcze ułatwiło analizę zachowań konsumentów.

Na początku ery zakupów internetowych zespół Dunn i Humby’ego wprowadził zakładkę pod nazwą „Moje ulubione”, w której wszystkie produkty zakupione z kartą klubową wyróżniano natychmiast po zalogowaniu się klienta na stronie Tesco. Sukces nowej funkcjonalności był porównywalny z sukcesem samej karty. Klienci mogli szybko zlokalizować poszukiwane produkty bez konieczności nawigowania przez kolejne ekrany. Sprzedaż wzrosła, a klienci byli zadowoleni.

Były jednak wyjątki. Wkrótce po wprowadzeniu nowej zakładki na stronie Tesco pewna klientka poskarżyła się firmie, że jej dane zostały sfałszowane. Kiedy zamawiała produkty przez internet, na jej liście ulubionych produktów wyświetliły się prezerwatywy. Nie mógł ich kupić jej mąż, bo ich nie używał. Na prośbę klientki analitycy Tesco przejrzeli historię karty i ustalili, że wszystko się zgadza. Firma wolała jednak nie doprowadzać do kryzysu małżeńskiego i podjęła dyplomatyczną decyzję – przeprosiła klientkę za „błąd w systemie” i usunęła inkryminowane artykuły z jej listy ulubionych produktów.

Clive Humby w swojej książce o Tesco pisze, że jest to żelazne, choć niepisane prawo w firmie: kiedy tylko na liście zakupowej pojawiają się produkty potencjalnie kłopotliwe, Tesco przeprasza i usuwa dane.

W podobnym duchu wypowiedział się Eric Schmidt, kiedy był prezesem zarządu firmy Google – że stara się zawsze wyczuwać, gdzie przebiega granica. „Polityką Google'a jest zbliżać się do śliskiej granicy, ale jej nie przekraczać”[4].

Kiedy jednak zgromadzisz wystarczająco dużą liczbę informacji, trudno powiedzieć, czy w ogóle zostanie jeszcze coś, czego nie zdołasz dowiedzieć się o człowieku. Produkty spożywcze mówią nie tylko o tym, co lubisz jeść, lecz także – jaką jesteś osobą. Wystarczy dokładnie przeanalizować nawyki konsumenckie klienta, żeby dobrze poznać jego charakter. Czasem – jak w przypadku wspomnianych wyżej prezerwatyw – dowiesz się o sprawach, o których wolałabyś nie wiedzieć. Częściej jednak gdzieś głęboko w gąszczu naszych danych można znaleźć perełki, które firmy potrafią zamienić w złoto.

Rynek Targetu

W 2002 roku amerykańska sieć wielkich dyskontów Target zaczęła szukać w danych swoich klientów nieoczywistych prawidłowości[5]. Target sprzedaje wszystko – od mleka i bananów, przez pluszowe zabawki, po meble ogrodowe – i jak chyba każda sieć detaliczna od początku obecnego stulecia, ma sposoby na to, żeby za pomocą numerów kart kredytowych i wyników ankiet przypisać klienta do wszystkiego, co ten kiedykolwiek kupił w jego sklepach, a co za tym idzie – dokładnie przeanalizować zachowania swoich konsumentów.

Historia ta wywołała skandal w całej Ameryce. Otóż najpierw analitycy Targetu odkryli, że wyraźny wzrost częstotliwości kupowania przez kobiety nieperfumowanych balsamów do ciała dla niemowląt jest często poprzedzony utworzeniem w sklepie przez te same klientki listy prezentów na *baby shower*. Dane wskazywały na wyraźne trendy. Kobiety na początku drugiego trymestru ciąży zaczynają się martwić o rozstępy, a więc wzrost zainteresowania kremami nawilżającymi podpowiadał prawdopodobny scenariusz rozwoju wypadków. Zaglądasz w dane sprzed paru miesięcy i widzisz, że te same kobiety wywożą z supermarketów pełne kosze witamin i suplementów takich jak wapń i cynk. Skaczesz więc znowu kilka miesięcy do przodu i dane mówią ci nawet, kiedy kobieta ma

termin, bo tuż przed rozwiązaniem w koszyku lądują największe opakowania waty[6].

Przyszła matka to marzenie każdej sieci. Zrób z niej lojalną klientkę, dopóki jest w ciąży, a prawie na pewno kobieta będzie kupować twoje produkty jeszcze długo po urodzeniu dziecka; zachowania konsumentek kształtują się bardzo szybko pod wpływem głodnych niemowląt, które zaczynają się głośno dopominać o swoje w trakcie cotygodniowej wyprawy mamy do supermarketu. Informacje zgromadzone przez analityków Targetu miały nieocenioną wartość, bo mogły zapewnić sieci zdecydowaną przewagę nad konkurencją w branży.

Dalej poszło już łatwo. Na bazie danych swoich klientek Target uruchomił algorytm obliczający prawdopodobieństwo zajścia w ciążę. Kobietom, które przekroczyły określony próg, automatycznie wysyłano pocztą serię kuponów na produkty mogące zainteresować przyszłe matki: pieluchy, balsamy do ciała, chusteczki nawilżane i tak dalej.

Z początku nie działo się nic strasznego, ale mniej więcej rok później do jednego z supermarketów Targetu w Minneapolis wparował ojciec nastolatki, żądając natychmiastowej rozmowy z kierownikiem sklepu. Mężczyzna się wściekł, bo jego córka dostała pocztą przesyłkę z kuponami na produkty ciążowe – czyżby Target zachęcał młodzież do uprawiania niebezpiecznego seksu? Kierownik sklepu najmocniej go przepraszał i kilka dni później zatelefonował jeszcze do niego, żeby przekazać mu od firmy wyrazy ubolewania. Wtedy już jednak, jak można przeczytać w artykule w „New York Timesie", ojciec nastolatki sam był gotów przeprosić kierownika sklepu.

„Odbyłem rozmowę z córką – powiedział mu – i okazuje się, że w moim domu dzieją się rzeczy, o których kompletnie nie miałem pojęcia. Córka ma termin w sierpniu".

Nie wiem, co ty o tym sądzisz, ale dla mnie cała ta sytuacja – kiedy marketowy algorytm informuje ojca o ciąży jego nastoletniej córki, zanim dziewczyna będzie gotowa, żeby sama mu o tym powiedzieć – to już stanowczo przekroczenie granicy. Ten przypadek nie wystarczył jednak Targetowi, by całkiem zrezygnować ze swojego algorytmu.

Wyjaśnił to jeden z dyrektorów sieci: „Zaobserwowaliśmy, że jeżeli kobiety w ciąży nie wiedzą o tym, że są monitorowane, wykorzystują nasze kupony. Zakładają po prostu, że wszystkie klientki sklepu dostały tę samą przesyłkę z kuponami na pieluchy i łóżeczko. Jeśli tylko ich nie spłoszymy, wszystko jest OK".

A zatem Target ciągle używa po cichu swojego algorytmu ciążowego, podobnie jak większość współczesnych sieci. Z tą różnicą, że teraz do kuponów typowo ciążowych dodaje kupony na bardziej neutralne produkty, żeby klientki się nie zorientowały, iż są profilowane: obok reklamy łóżeczka dla dziecka może się pojawić reklama kieliszków do wina. Albo przy kuponie na ubranka niemowlęce znajdzie się reklama wody kolońskiej.

Target nie jest jedyną siecią, która korzysta z podobnych narzędzi. Historie o tym, co można wyczytać z twoich danych, rzadko przedostają się do prasy, ale algorytmy do ich analizowania istnieją – dyskretnie ukryte głęboko na tyłach korporacyjnego frontu. Mniej więcej rok temu rozmawiałam z dyrektorem do spraw danych w pewnej firmie ubezpieczeniowej. Otóż firma ta korzysta ze szczegółowego profilu nawyków zakupowych pozyskanego z programu lojalnościowego sieci marketów. Analiza tych danych wykazała, że osoby gotujące w domu rzadziej zgłaszają szkody mieszkaniowe, a więc są dla ubezpieczyciela bardziej dochodowe. To oczywiście zgodne z tym, co nam podpowiada zdrowy rozsądek – jest raczej oczywiste, że grupa ludzi gotowych inwestować swój czas, siłę i pieniądze w przyrządzenie od zera wymyślnej potrawy we własnej kuchni musi się z grubsza pokrywać z grupą osób, które nie pozwalają dzieciom grać w domu w piłkę. Tylko skąd firma ubezpieczeniowa wie, którzy jej klienci lubią gotować w domu? Otóż w koszykach tych klientów zidentyfikowano kilka produktów, które korelują z niskim „wskaźnikiem szkodowości". Jak powiedział mój rozmówca, najważniejszym z tych produktów, który pozwoli na najpewniejszą identyfikację osoby odpowiedzialnej i dumnej z samodzielnego prowadzenia domu, jest świeży koper włoski.

Jeśli aż tyle informacji da się wyciągnąć z nawyków zakupowych osób funkcjonujących w świecie rzeczywistym, to można sobie wyobrazić, jakie wnioski dałoby się wyciągnąć, gdyby uzyskało się dostęp do pełniejszych

danych. I czego można by się dowiedzieć o drugim człowieku, gdyby rejestrowało się wszystkie jego poczynania w świecie wirtualnym.

Dziki Zachód

Palantir Technologies należy do najbardziej udanych start-upów w całej historii Doliny Krzemowej. Wartość firmy założonej w 2003 roku przez Petera Thiela (znanego założyciela Pay-Pala) dziś szacuje się na astronomiczną kwotę 20 miliardów dolarów[7]. To mniej więcej tyle, ile warty jest na rynku Twitter, choć jest bardzo prawdopodobne, że o Palantirze nigdy nie słyszeliście. On natomiast – i możesz mi wierzyć, bo wiem, co mówię – z całą pewnością słyszał o tobie.

Palantir to zaledwie jeden z przedstawicieli nowego pokolenia firm określanych zbiorczym mianem „brokerów danych"; kupują one i gromadzą osobiste informacje o użytkownikach sieci, żeby z zyskiem odsprzedać je lub udostępnić podmiotom trzecim. Jest ich znacznie więcej: Acxiom, Corelogic, Datalogix, e-Bureau – cały rój gigantów, z którymi najprawdopodobniej nie mieliście nigdy bezpośredniej styczności, mimo że jesteśmy bezustannie przez nie obserwowani, a nasze zachowania dokładnie analizowane[8].

Za każdym razem, kiedy kupujesz coś w internecie, kiedy zapisujesz się do newslettera albo rejestrujesz na takiej czy innej stronie, szukasz informacji o nowym samochodzie albo wypełniasz kartę gwarancyjną online, kupujesz nowe mieszkanie albo rejestrujesz się przed wyborami – za każdym razem, kiedy zostawiasz w sieci jakikolwiek osobisty ślad, twoje dane są gromadzone i sprzedawane brokerowi. Pamiętasz, kiedy powiedziałaś agentowi z biura nieruchomości, jakiego domu szukasz? Sprzedane brokerowi. Albo te informacje, które wpisałeś kiedyś w porównywarce ubezpieczeń? Sprzedane brokerowi. W pewnych przypadkach nawet cała twoja historia przeglądania stron internetowych może zostać spakowana i odsprzedana podmiotom trzecim[9].

Zadaniem brokera jest połączenie wszystkich zgromadzonych i zakupionych danych, powiązanie ze sobą najróżniejszych informacji i skompilowanie z nich szczegółowego profilu twojego elektronicznego

cienia. Nie ma w tym absolutnie żadnej przesady – w niektórych bazach brokerów danych mogłabyś odszukać plik oznaczony niepowtarzalnym numerem identyfikacyjnym (którego jednak nigdy nie poznasz) i znaleźć w nim wszelkie ślady swojej aktywności w świecie wirtualnym. Twoje imię i nazwisko, data urodzenia, wyznanie, typowy sposób spędzania wakacji, wykorzystanie karty kredytowej, twoja wartość netto, waga, wzrost, przekonania polityczne, skłonności do hazardu, stopień i rodzaj niepełnosprawności, zażywane leki, czy przerywałaś ciążę, czy twoi rodzice się rozwiedli, czy masz skłonności do nałogów, czy zostałaś zgwałcona, co sądzisz o powszechnym dostępie do broni palnej, jaka jest twoja przewidywana i rzeczywista orientacja seksualna i czy jesteś łatwowierny. Mówimy tu o tysiącach szczegółowych informacji, pogrupowanych w tysiącach kategorii, przechowywanych gdzieś w mrocznych zakamarkach sieci na ukrytych serwerach – informacji o dosłownie każdym z nas[10].

Podobnie jak z algorytmem ciążowym Targetu, większość tych informacji to wynik wnioskowania. Subskrypcja czasopisma „Wired" może implikować, że interesujesz się techniką; licencja na broń palną może oznaczać, że lubisz polować. Brokerzy na każdym kroku korzystają z inteligentnych, choć prostych algorytmów do wzbogacania swoich danych. Dokładnie to samo robią supermarkety, tylko nie na masową skalę.

Zyski są ogromne. Dzięki znajomości naszych nawyków brokerzy danych uniemożliwiają oszustom podszywanie się pod nas w internecie. Poznanie naszych preferencji skutkuje tym, że reklamy, na które natrafiamy podczas surfowania po sieci, są dobrze dopasowane do naszych zainteresowań i potrzeb. A to na pewno znacznie uprzyjemnia użytkowanie sieci, bo kto chciałby być dzień w dzień bombardowany zupełnie go nieinteresującymi reklamami prawników medycznych czy ubezpieczeń kredytu. A ponieważ reklamy mogą być dostarczane bezpośrednio do potencjalnych klientów, ogólne koszty marketingu maleją, dzięki czemu małe przedsiębiorstwa oferujące świetne produkty mogą pozyskiwać nowych nabywców – na czym zyskujemy wszyscy.

Niemniej z pewnością myślisz już o tym samym co ja – mamy tu również całe spektrum problemów, które mogą się pojawić, kiedy komputery zaczną postrzegać nas, ludzi, jako ciąg kategorii. Wrócę do tego za chwilę,

najpierw jednak chciałabym wytłumaczyć, jak to się dzieje, że surfując sobie beztrosko po internecie, ni stąd, ni z zowąd natrafiasz na reklamę internetową – i jaką rolę odgrywa w tym procesie broker danych.

Wyobraźmy sobie, że jestem posiadaczką luksusowego biura podróży, o wyszukanej nazwie Fry's. Od wielu lat namawiam ludzi, żeby rejestrowali się na mojej stronie, i teraz mam pewną listę adresów e-mailowych. Gdybym chciała dowiedzieć się czegoś więcej o tych ludziach – na przykład o jakich wakacjach marzą – mogłabym przesłać tę listę do brokera danych, który przefiltrowałby swoje bazy i odesłał mi ją wzbogaconą o interesujące mnie informacje. To tak jakby rozszerzyć arkusz kalkulacyjny o dodatkową kolumnę. Kiedy teraz odwiedzasz moją stronę w internecie, ja już wiem, że masz szczególne upodobanie do wysp tropikalnych, więc częstuję cię reklamą hawajskiego raju.

To jest opcja pierwsza. W opcji drugiej wyobraźmy sobie, że biuro podróży Fry's wydzieliło mały fragment swojej strony internetowej i zamierza udostępnić go innym reklamodawcom. Znów kontaktuję się z brokerem danych i przekazuję mu zgromadzone przez siebie informacje o użytkownikach mojej strony. Broker szuka firm, które chciałyby się u mnie reklamować. Na potrzeby naszej historii wyobraźmy sobie teraz, że broker znalazł dla mnie potencjalnego reklamodawcę – firmę sprzedającą kremy z filtrem UV. Chcąc ją przekonać, że moje biuro podróży ma klientów, którzy mogliby kupować jej produkty, broker pokazałby jej niektóre z wydedukowanych cech użytkowników strony biura podróży Fry's: na przykład odsetek osób mających rude włosy. Mogłoby też być tak, że firma sprzedająca kremy z filtrem przekazałaby brokerowi listę adresów e-mailowych swoich klientów, żeby broker sam wyliczył, w jakim stopniu obie grupy do siebie pasują. I jeżeli sprzedawcy kremów z filtrem się zgodzą, na mojej stronie pojawi się ich reklama – a ja i broker dostaniemy za to pieniądze.

Na razie metody te nie różnią się znacznie od technik od zawsze stosowanych przez działy marketingu. Istnieje jednak jeszcze opcja trzecia, w której – jak dla mnie – sprawy znów zaczynają się ocierać o śliską granicę. W tym wariancie biuro podróży Fry's szuka nowych klientów; interesują mnie mężczyźni i kobiety w wieku powyżej 65 lat, którzy lubią wyspy tropikalne, a pieniądze nie są dla nich problemem, bo bardzo liczę

na to, że będą chcieli zakupić jedną z moich nowych luksusowych wycieczek po Karaibach. Cóż więc robię? Udaję się do brokera danych, który przeszukuje swoje bazy i sporządza dla mnie listę osób pasujących do mojego profilu idealnego klienta.

No i wyobraźmy sobie, że znalazłeś się na tej liście. Broker nigdy nie ujawni mi twoich personaliów, dowie się jednak, na których stronach bywasz regularnie. Bardzo prawdopodobne, że będzie miał dobre układy z jedną z twoich ulubionych witryn – może to być strona mediów społecznościowych albo portal informacyjny – i kiedy tylko, niczego nie podejrzewając, zalogujesz się na niej, broker zostanie o tym powiadomiony i dosłownie w tej samej chwili zareaguje umieszczeniem w twoim komputerze maleńkiego znacznika, znanego powszechnie jako *cookie*, czyli ciasteczko[6*]. Ciasteczko rozsyła do rozmaitych witryn w całym internecie sygnał mówiący o tym, że właśnie teraz należy ci podsunąć pod nos reklamę luksusowej wycieczki po Karaibach oferowanej przez biuro Fry's. Czy tego chcesz, czy nie, takie reklamy będą ci towarzyszyć na każdym twoim kroku w sieci.

Tu natrafiamy na pierwszy poważny problem. Bo co jeśli nie masz ochoty widzieć tej reklamy? Oczywiście bombardowanie wizjami wycieczki po Karaibach może być źródłem co najwyżej drobnego dyskomfortu, istnieją jednak reklamy, które mogą wpływać na psychikę człowieka na znacznie głębszym poziomie.

Kiedy Heidi Waterhouse straciła długo oczekiwaną ciążę[11], anulowała subskrypcje wszystkich cotygodniowych e-maili, które informowały ją o tym, jak się rozwija jej dziecko i jaki owoc odpowiada teraz wielkości płodu. Wypisała się ze wszystkich list mailingowych i prezentowych, na które się zapisała, nie mogąc się już doczekać daty rozwiązania. Niemniej – jak powiedziała na konferencji deweloperów w 2018 roku – nie było na Ziemi takiej siły, która zdołałaby uwolnić ją od reklam ciążowych atakujących ją w sieci dosłownie wszędzie. Jej cyfrowy cień żył w internecie własnym życiem, bez matki i dziecka. „Ludzie, którzy stworzyli ten system, nie wzięli takiej ewentualności w ogóle pod uwagę", powiedziała.

Czy to przez ludzką nieuwagę, czy wskutek celowego działania system rzeczywiście nader chętnie wyzyskuje luki w zabezpieczeniach. Firmy udzielające krótkoterminowych pożyczek mogą go wykorzystywać do

identyfikowania klientów ze złą historią kredytową; reklamy zakładów sportowych mogą być kierowane bezpośrednio do regularnych użytkowników portali hazardowych. Istnieją też obawy, że tego rodzaju profilowanie można wykorzystywać przeciw obywatelom, bo przecież na przykład entuzjaści sportów motocyklowych mogą zostać uznani za osoby mające bardzo niebezpieczne hobby, a miłośnicy bezcukrowych słodyczy zostaną zaszufladkowani jako diabetycy i w rezultacie towarzystwa ubezpieczeniowe odmówią im polisy. Analiza z 2015 roku wykazała, że Google podsuwa kobietom znacznie mniej reklam ofert pracy na najwyżej opłacanych stanowiskach dyrektorskich niż mężczyznom[12]. Kiedy pewnego razu afroamerykańska profesor Harvardu zauważyła, że wpisanie jej nazwiska do okienka przeglądarki skutkuje tym, iż Google zaczyna jej serwować reklamy skierowane do osób z przeszłością kryminalną (co doprowadziło do tego, że profesor Harvardu musiała udowodnić swojemu potencjalnemu pracodawcy, iż nigdy nie miała do czynienia z policją), postanowiła przeanalizować zawartość reklam kierowanych w sieci do przedstawicieli różnych grup etnicznych. I wyszło jej, że prawdopodobieństwo powiązania wyników wyszukiwania z ogłoszeniami zawierającymi słowa z rdzeniem „areszt" (na przykład „Czy byłeś kiedyś aresztowany?") jest nieproporcjonalnie wyższe dla nazwisk „czarnych" niż dla nazwisk „białych"[13].

Stosowanie tych metod nie ogranicza się do brokerów danych. Nie ma prawie różnicy między sposobem ich działania a funkcjonowaniem Google'a, Facebooka, Instagramu i Twittera; te internetowe mocarstwa nie zarabiają na kontach swoich użytkowników – ich modele biznesowe są rozwinięciem idei mikrotargetingu. Firmy te to w rzeczywistości gigantyczne silniki dystrybucji reklam, które zarabiają ogromne pieniądze na tym, że użytkownicy stworzonych przez nie portali społecznościowych aktywnie z nich korzystają, co chwila w coś klikając, żeby przeczytać sponsorowany post albo obejrzeć sponsorowany film czy sponsorowane zdjęcie. Bez względu na to, gdzie się w internecie obracasz, masz pewność, że w głębokim cieniu czają się dyskretne algorytmy, które bez twojej wiedzy i zgody operują informacjami o tobie. Algorytmy, dla których twoje najbardziej osobiste, intymne sekrety są wyłącznie towarem na sprzedaż.

Niestety, prawo wielu krajów nie za bardzo cię chroni. Działalność brokerów danych pozostaje w większej części zupełnie nieuregulowana i – szczególnie w Stanach Zjednoczonych – państwo raz za razem przepuszcza dobre okazje, żeby ograniczyć ich władzę. Na przykład w marcu 2017 roku amerykański Senat zagłosował za zniesieniem regulacji, które zakazywałyby brokerom danych sprzedawania twojej historii przeglądania bez twojej wyraźnej zgody. Kilka miesięcy wcześniej, w październiku 2016 roku, regulacje te zostały zatwierdzone przez Federalną Komisję Komunikacji, lecz po zmianie administracji pod koniec tego samego roku nowa republikańska większość w Komisji i republikański Kongres nie zgodziły się na ich wprowadzenie w życie[14].

Co to wszystko oznacza dla twojej prywatności? Pozwól, że zamiast odpowiedzieć od razu, przedstawię historię badań przeprowadzonych przez niemiecką dziennikarkę Sveę Eckert i analityka Andreasa Dewesa – to wystarczająco naświetli problem[15].

Eckert i jej zespół podszyli się pod brokera danych i wykorzystali przykrywkę do wykupienia historii przeglądania trzech milionów anonimowych obywateli Niemiec. (Zdobycie tych danych było łatwe. Mnóstwo firm oferuje nieprzebrane bogactwo tego rodzaju informacji o klientach brytyjskich czy amerykańskich – jedynym wyzwaniem było znalezienie danych niemieckich). Same informacje zostały zebrane przez jedno z rozszerzeń przeglądarki Google Chrome, które użytkownicy zainstalowali dobrowolnie, nie mając pojęcia o tym, że wtyczka będzie śledzić ich poczynania w sieci[7*].

Tak Eckert weszła w posiadanie wprost monstrualnie wielkiej bazy adresów URL. Była to lista wszystkich miejsc, które użytkownicy internetu odwiedzili w ciągu miesiąca. Każde wyszukanie, każda strona, każde kliknięcie. I wszystko to było legalnie wystawione na sprzedaż.

Jedyną trudność mogła zespołowi Eckert sprawić anonimowość danych. To pewnie byłaby dobra wiadomość dla tych wszystkich Niemców, których historie przeglądania sprzedano brokerowi, prawda? Oszczędziłoby im to zapewne sporo wstydu. Ale nic z tego! Podczas prezentacji na hakerskiej konferencji DEFCON w 2017 roku członkowie zespołu wykazali, że

deanonimizacja gigantycznych baz historii przeglądania jest śmiesznie prosta.

Niekiedy tożsamość użytkownika dało się wyczytać bezpośrednio z samych adresów URL. Wystarczy, że ktoś odwiedził stronę Xing.com, niemiecki odpowiednik platformy LinkedIn. Kiedy klikniesz na swoje zdjęcie profilowe na Xingu, zostaniesz przekierowany na stronę z adresem, który będzie wyglądał mniej więcej tak:

www.xing.com/profile/Hannah_Fry?sc_omxb_p

Nazwisko w adresie strony demaskuje mnie natychmiast, a tekst umieszczony po nim oznacza, że użytkownik jest zalogowany i przegląda własny profil; dzięki temu Eckert i jej ludzie mogli jednoznacznie powiązać stronę z użytkownikiem. Podobnie było z Twitterem – wystarczyło zajrzeć na stronę analityki swojego profilu, żeby ujawnić swoją tożsamość. Na potrzeby użytkowników, których dane nie zawierały tak oczywistego nawiązania do tożsamości jak imię i nazwisko, zespół Eckert miał w rękawie inny trik. Każdy, kto zamieścił online jakikolwiek post (tweet o stronie internetowej albo udostępnienie swojej listy odtwarzania z YouTube'a), jakikolwiek publiczny odcisk danych powiązanych ze swoim prawdziwym imieniem i nazwiskiem – automatycznie wychodził z cienia anonimowości. Zespół Eckert zastosował prosty algorytm do korelowania tożsamości publicznych z danymi zanonimizowanymi[16]: filtrował on listę adresów URL i wyszukiwał w bazie danych osoby odwiedzające strony internetowe, na których zamieszczono anonimowe linki, dokładnie tego samego dnia o tej samej porze. W ten sposób zespół zdołał zidentyfikować praktycznie wszystkich użytkowników ze swojej bazy danych – i tak uzyskał dostęp do pełnej miesięcznej historii ich poczynań w sieci.

W tej bazie trzech milionów obywateli niemieckich znalazło się wiele osób na wysokich stanowiskach. Między innymi polityk, który szukał w internecie pewnego leku. I oficer policji, który skopiował i wkleił do automatycznego translatora firmy Google dokument zawierający dane wrażliwe – wszystkie informacje o tym dokumencie znalazły się w adresie URL i w rezultacie stały się widoczne dla zespołu Eckert. Był też pewien

sędzia, którego historia przeglądania nosiła ślady codziennych wizyt w dość szczególnej strefie internetu. Oto mały wybór stron odwiedzonych przez niego w ciągu ośmiu minut w sierpniu 2016 roku:

18.22: http://www.tubegalore.com/video/amature-pov-ex-wife-in-leather-pants-gets-creampie42945.html

18.23: http://www.xxkingtube.com/video/pov_wifey_on_sex_stool_with_beaded_thong_gets_creampie_4814.html

18.24: http://de.xhamster.com/movies/924590/office_lady_in_pants_rubbing_riding_best_of_anlife.html

18.27: http://www.tubegalore.com/young_tube/5762-1/page0

18.30: http://www.keezmovies.com/video/sexy-dominatrix-milks-him-dry-1007114?utm_sources

Pomiędzy tymi codziennymi sesjami sędzia rutynowo poszukiwał w sieci imion dziecięcych, wózków dla niemowląt oraz szpitali położniczych. Zespół Eckert wywnioskował, że jego partnerka w tym czasie spodziewała się dziecka.

Musimy powiedzieć to sobie jasno i wyraźnie: sędzia nie robił niczego zakazanego prawem. Wiele osób – w tym ja – broniłoby poglądu, że nie robił w ogóle niczego złego. Niemniej podobne informacje o nim byłyby idealnym narzędziem w rękach osób, które nie zawahałyby się pana sędziego zaszantażować czy narobić mu wstydu przed jego własną rodziną.

I tutaj już wykraczamy bardzo daleko poza śliską granicę (Google'a): kiedy ktoś bez twojej wiedzy gromadzi intymne, wrażliwe informacje o tobie, żeby potem za ich pomocą móc tobą manipulować. Właśnie coś takiego przydarzyło się osobom, które wzięła na celownik brytyjska firma konsultingu politycznego o nazwie Cambridge Analytica.

Cambridge Analytica

Znasz już tę historię, przynajmniej w ogólnych zarysach.

Od lat osiemdziesiątych ubiegłego wieku psychologowie kwantyfikują osobowość człowieka, przyznając określoną liczbę punktów za następujące cechy: otwartość na doświadczenie, sumienność, ekstrawertyczność, ugodowość i neurotyczność. To standardowy, użyteczny sposób opisania, jaką jesteś osobą.

W roku 2012, czyli rok przed powstaniem spółki Cambridge Analytica, grupa naukowców z uniwersytetów Cambridge i Stanforda postanowiła sprawdzić, czy da się powiązać tych pięć cech osobowości ze stronami, które ludzie „lajkują" na Facebooku[17]. Aby odnaleźć potencjalne powiązania prawdziwego charakteru badanych z ich tożsamościami internetowymi, naukowcy stworzyli na Facebooku quiz umożliwiający użytkownikom przeprowadzenie prawdziwego testu psychometrycznego. Osoby, które ściągnęły quiz, miały świadomość, że przekazują jego twórcom zarówno swoją historię polubień na Facebooku, jak i – za pośrednictwem zawartych w nim pytań – swoją rzeczywistą psychometrię.

Łatwo sobie wyobrazić, w jaki sposób „lajki" mogą się łączyć z osobowością człowieka. W artykule opublikowanym w 2013 roku[18] członkowie zespołu napisali, że użytkownicy, którzy lubią Salvadora Dalego, medytację czy TED Talks, niemal na pewno będą mieli wysoki wynik w zakresie otwartości na nowe doświadczenia. Z kolei osoby, które lubią imprezy, taniec i Snooki z reality show *Ekipa z New Jersey*, są z reguły bardziej ekstrawertyczne. Badania zakończyły się sukcesem. Dzięki odkrytym powiązaniom zespół mógł stworzyć algorytm, który potrafił wnioskować o cechach osobowości, opierając się wyłącznie na facebookowych „lajkach".

W 2015 roku ci sami badacze opublikowali drugi artykuł[19], w którym twierdzili, że trzysta „lajków" z profilu facebookowego dowolnej osoby wystarczy, żeby algorytm ocenił jej charakter dokładniej niż jej mąż czy żona.

Przeskoczmy szybko w czasie do dnia dzisiejszego. Oto nasza uniwersytecka grupa badawcza – Ośrodek Psychometrii na uniwersytecie w Cambridge – poszerza możliwości swojego algorytmu: teraz może on prognozować, jakim jesteś człowiekiem, na podstawie treści z twojego profilu na Tweeterze. Ośrodek ma własną stronę internetową, otwartą dla

każdego, na której można samemu się przekonać, czy to nie czcza przechwałka. Ponieważ moje konto na Tweeterze i tak jest publiczne, postanowiłam sprawdzić, jak to działa: załadowałam moją historię tweetów i wypełniłam tradycyjny kwestionariusz, żeby porównać jego rezultaty z przewidywaniami algorytmu. Komputer zdołał ocenić mnie poprawnie w trzech (z pięciu standardowych) wymiarach osobowości. Tradycyjny test osobowości wykazał natomiast, że jestem znacznie bardziej ekstrawertyczna i neurotyczna, niż można by wnioskować na podstawie treści mojego profilu na Tweeterze[8*].

Wszystkie działania zespołu zmierzały do zastosowania nowego algorytmu w marketingu. Dlatego też w 2017 roku[20] ten sam zespół badaczy zaczął eksperymentować z wysyłaniem reklam dopasowanych do indywidualnych miar osobowościowych: rozsyłał na Facebooku reklamy kosmetyku ekstrawertykom, opatrując je hasłem: „Tańcz, jakby nikt nie patrzył (choć wszyscy wybałuszają oczy)", podczas gdy introwertycy widzieli obraz uśmiechniętej dziewczyny stojącej przed lustrem, podpisany hasłem: „Piękno nie musi krzyczeć".

W równoległym eksperymencie grupie docelowej z wysoką punktacją w skali otwartości na doświadczenia pokazywano reklamy krzyżówek z hasłem: „Arystoteles? Seszele? Uruchom kreatywność i rzuć wyzwanie wyobraźni! Nieograniczona liczba krzyżówek!". Te same łamigłówki reklamowano osobom o niskim wyniku otwartości, tym razem korzystając z hasła: „Zasiądź do swojej ulubionej rozrywki – krzyżówki, która testuje szaradzistów od pokoleń!". Badacze chwalili się, że dopasowanie reklamy do charakteru osoby przełożyło się na 40 procent więcej kliknięć i aż 50 procent więcej zakupów niż przy zastosowaniu ogólnej, niezindywidualizowanej reklamy. Na reklamodawcach te liczby robią ogromne wrażenie.

Kiedy badacze z Cambridge publikowali wyniki swoich badań, inni już stosowali ich metody w praktyce – w tym spółka Cambridge Analytica, która miała je wykorzystywać podczas kampanii wyborczej Donalda Trumpa.

Cofnijmy się teraz trochę w czasie. Nie ma większych wątpliwości co do tego, że Cambridge Analytica korzystała z identycznych technik jak moja wymyślona luksusowa agencja podróży Fry's: tworzyła niewielkie zbiory

osób, które uznała za podatne na sugestie, po czym kierowała reklamy bezpośrednio do nich, rezygnując z reklamy masowej. Jednym z odkryć Cambridge Analytica było znalezienie silnego związku między zbiorem osób jeżdżących starymi, dobrymi amerykańskimi fordami i zbiorem zarejestrowanych zwolenników Partii Republikańskiej. Następnym krokiem było odszukanie miłośników fordów, którzy oficjalnie nie tworzyli elektoratu tej partii, i sprawdzenie, czy da się na nich wpłynąć proamerykańskimi reklamami, które dobrze wpisują się w patriotyczne ciągoty republikanów. W pewnym sensie było to powielenie elementu kampanii wyborczej, w którym kandydaci jadą do miejscowości zamieszkanych przez niezdecydowanych wyborców, żeby chodzić od drzwi do drzwi i osobiście przekonywać ich do głosowania na siebie. Z kolei na poziomie sieci praktyka ta niczym się nie różni od tego, co robili Obama i Clinton podczas swoich kampanii. Każda większa partia w świecie zachodnim korzysta z pogłębionej analityki i wyborczego mikrotargetingu.

Jeżeli jednak możemy wierzyć zapisowi z ukrytej kamery kanału Four News, firma Cambridge Analytica wykorzystywała również portrety psychologiczne użytkowników internetu do grania na politycznych emocjach republikańskiego elektoratu – na przykład odszukiwała samotne matki o wysokim stopniu neurotyczności, żeby żerując na ich strachu przed napaścią na swój dom, przeciągnąć je na stronę zwolenników powszechnego dostępu do broni palnej. Marketing komercyjny z pewnością stosuje te techniki na masową skalę i z dużym prawdopodobieństwem można przyjąć, że używano ich również w innych kampaniach politycznych.

Ale to nie koniec. Spółkę Cambridge Analytica oskarża się także o tworzenie reklam upozorowanych na materiał dziennikarski. Pewien informator powiedział „Guardianowi", że jedną z najbardziej skutecznych reklam kampanii Trumpa była interaktywna grafika zatytułowana „10 niewygodnych faktów o Clinton Foundation"[21]. Inny demaskator powiedział jeszcze więcej – że „artykuły" podsuwane przez Cambridge Analytica często zawierały stwierdzenia jawnie fałszywe[22].

Na potrzeby dalszego wywodu przyjmijmy, że wszystko, co zostało wyżej powiedziane, jest prawdą, czyli że firma konsultingowa Cambridge

Analytica umieszczała na Facebooku fałszywe „wiadomości", próbując zmanipulować nimi użytkowników o znanym sobie profilu psychologicznym. I teraz zadajmy sobie pytanie: czy to się mogło udać?

Mikromanipulacja

Istnieje pewna asymetria w postrzeganiu możliwości personalizowanych reklam politycznych. Lubimy uważać się za osoby niezależne i odporne na manipulację, a jednocześnie wyobrażamy sobie, że inni – a już zwłaszcza osoby o odmiennych poglądach politycznych – są do bólu naiwni. Rzeczywistość będzie raczej wypadkową obu tych skrajności.

Bardzo dobrze wiemy, że posty, które widzimy na Facebooku, mogą wpływać na nasze emocje. W kontrowersyjnym eksperymencie z 2013 roku pracownicy Facebooka zmanipulowali posty ze źródeł zasubskrybowanych przez 689 003 użytkowników portalu bez ich wiedzy (i zgody), żeby sprawdzić, czy da się wpłynąć na ich emocje i nastroje[23]. Eksperymentatorzy zablokowali wszystkie docierające do tych osób posty zawierające słowa o znaczeniu pozytywnym, a następnie zrobili to samo z postami zawierającymi słowa o znaczeniu negatywnym, za każdym razem obserwując reakcje użytkowników. Osoby, które widziały mniej negatywnych treści na swoim profilu, zaczęły same publikować więcej postów o pozytywnym wydźwięku. Natomiast te osoby, którym blokowano treści pozytywne, same zaczęły w swoich postach używać więcej słów o znaczeniu negatywnym. Wniosek: nawet jeśli sądzimy, że jesteśmy odporni na manipulację emocjonalną, prawdopodobnie jesteśmy w błędzie.

Z eksperymentu Epsteina opisanego w rozdziale *Władza* wiemy również, że nawet odpowiednie uszeregowanie stron z wynikami wyszukiwania może wystarczyć, żeby dotąd niezdecydowany wyborca polubił faworyzowanego przez przeglądarkę kandydata. Wiemy też – z prac badaczy szacownych uniwersytetów, których algorytmy zostały później przeflancowane na grunt polityki przez Cambridge Analytica – że reklamy są bardziej skuteczne, kiedy odwołują się do cech osobowości potencjalnych odbiorców.

Przykłady te stanowią solidne argumenty na poparcie tezy, że stosowanie wyżej opisanych metod może wpływać na to, na kogo ludzie zagłosują – podobnie jak wpływa na to, na co ludzie wydają pieniądze. Ale – i jest to „ale" naprawdę spore – zanim podejmiesz decyzję, kogo wybrać lub na co wydać pieniądze, musisz wiedzieć jeszcze jedną rzecz.

Wszystko, o czym mówiliśmy powyżej, jest prawdą, ale wiedz, że efekty tych działań są znikome. W eksperymencie pracowników Facebooka użytkownicy rzeczywiście skłaniali się ku temu, żeby publikować więcej postów o pozytywnym wydźwięku, kiedy zablokowano im dostęp do treści negatywnych. Tylko że różnica nie przekroczyła jednej dziesiątej procenta.

Podobnie było z personalizowanymi reklamami kosmetyków – ten sprzedawany introwertyczkom miał większe powodzenie, kiedy reklama uwzględniała charakter potencjalnej klientki, lecz efekty uzyskane dzięki tej metodzie były ledwo uchwytne. Normalna reklama nakłoniła do kliknięcia 31 osób z 1000. Reklama spersonalizowana – 35 z 1000. Nawet ten wspominany na stronach 59–60 50-procentowy wzrost sprzedaży – wypisany wielkimi literami w tytule pracy naukowej – oznaczał po prostu 16 kliknięć zamiast 11, z puli tysiąca.

Te metody mogą działać, to prawda. Ale reklamodawcy nie ładują przecież swoich reklam prosto do mózgów biernych odbiorców. Aż tak łatwym celem jeszcze nie jesteśmy – mamy znacznie lepsze zdolności ignorowania reklam i prześwietlania propagandy, niżby tego chciały firmy serwujące nam podprogowo takie czy inne treści. Skutek jest taki, że nawet te najlepsze, najbardziej przemyślnie sprofilowane kampanie reklamowe zdołają wpłynąć na odbiorców w bardzo niewielkim zakresie.

Ale w wyborach ten znikomy wpływ może się okazać, przynajmniej teoretycznie, języczkiem u wagi. W populacji złożonej z dziesiątek czy setek milionów ludzi zmiana poglądów u jednej osoby na tysiąc może się przełożyć na wcale pokaźną siłę. Kiedy przypomnisz sobie – na co wskazał w swoim artykule dla „Spectatora" Jamie Bartlett – że Trump wygrał w Pensylwanii różnicą 44 000 głosów, w Wisconsin 22 000, a w Michigan zaledwie 11 000, to niewykluczone, że efekt wspomnianych metod nie musi być większy – ten marginalny, niecały procent zupełnie wystarczy[24].

Prawda jest taka, że nie jesteśmy w stanie dzisiaj ocenić, jaki wpływ na wyniki wyborów prezydenckich w Stanach Zjednoczonych miały tego typu

zabiegi marketingowe. Nawet gdybyśmy mieli dostęp do wszystkich informacji, nie cofniemy się już w czasie i nie rozplączemy lepkiej sieci przyczyn i skutków, żeby wyłuskać z niej ten jeden jedyny powód takiej, a nie innej decyzji tego, a nie innego wyborcy. Co się stało, to się nie odstanie. Teraz liczy się tylko jedno: co nas czeka w przyszłości?

Oceń mnie

Pamiętajmy, że wszyscy skorzystaliśmy na takim modelu internetu. Ludzie na całym świecie mają swobodny i prosty dostęp do globalnych sieci natychmiastowej komunikacji, mogą czerpać z nieprzebranych zasobów ludzkiej wiedzy, co minutę aktualizowanych informacji ze wszystkich zakątków planety oraz rewelacyjnych programów i technologii stworzonych przez prywatne firmy i opłaconych przez reklamy. Taki był układ, na który wszyscy poszliśmy: darmowa technologia w zamian za nasze dane i możliwości ich wykorzystywania w celu wpływania na nas i czerpania z nas zysków. Wszystko, co najlepsze i najgorsze w kapitalizmie, w jednym prostym barterze.

Możemy uznać, że te warunki nas zadowalają. I to będzie absolutnie w porządku. Musimy jednak uzmysłowić sobie zagrożenia, jakie wiążą się z samym gromadzeniem informacji osobistych o użytkownikach sieci. Musimy się zastanowić, do czego może doprowadzić istnienie baz danych wrażliwych – i warto wybiec myślą dalej w przyszłość, poza zagadnienia prywatności i potencjalnych zagrożeń dla demokracji (jakby tego było nam mało), ponieważ ta dystopia ma jeszcze inny wymiar. Bogate, sprzężone ze sobą bazy informacji mogą mieć zastosowania, o których nam się nawet nie śniło, bardziej pasujące do scenariusza popularnego serialu Netfliksa *Czarne lustro* niż do naszej rzeczywistości. Właściwie to już mają. Przyjrzyjmy się systemowi oceny obywateli Sesame Credit wdrożonemu przez rząd chiński.

Wyobraź sobie, że każda pojedyncza informacja, którą broker danych zdołał o tobie zebrać, zostaje zamieniona na punkty. Będzie liczyć się wszystko: nie tylko rzeczy oczywiste, takie jak twoja historia kredytowa, numer twojego telefonu czy adres, lecz także twoje zachowanie na co dzień, twoje posty w mediach społecznościowych, dane z twojej aplikacji

do zamawiania taksówek, a nawet informacje wprowadzone do internetowego serwisu randkowego. System sumuje to wszystko i daje ci od 350 do 950 punktów.

Sesame Credit nie ujawnia szczegółów działania swojego „bardzo skomplikowanego" algorytmu punktacji, ale w wywiadzie dla pekińskiej Caixin Media dyrektor techniczny firmy, Li Yingyun, uchylił nieco rąbka tajemnicy: „Na przykład osoba, która przez dziesięć godzin dziennie gra w gry wideo, zostałaby uznana za lenia. A osoba, która często kupuje pieluchy, zostałaby z dużym prawdopodobieństwem uznana za odpowiedzialnego rodzica"[25].

Jeżeli mieszkasz w Chinach, punkty mają znaczenie. Kiedy twój wynik przekroczy 600 punktów, możesz dostać specjalną kartę kredytową. Przekrocz 666, a zostanie ci przyznany wyższy limit kredytowy. Obywatele, którzy uzyskają powyżej 650 punktów, mogą wynająć samochód bez wpłacania kaucji i korzystać na pekińskim lotnisku z terminalu dla VIP-ów. Każdy, kto uzbiera więcej niż 750 punktów, może się ubiegać o szybką wizę do Europy[26].

Wszystko to tymczasem jeszcze zabawa w żołnierzyki, bo przystąpienie do systemu jest dobrowolne. Kiedy jednak w roku 2020 stanie się obligatoryjne, nisko notowani obywatele zaczną to odczuwać we wszystkich dziedzinach życia. Publikacja rządowa o nowym systemie podaje przykłady kar, które będzie można wymierzać za domniemane nieposłuszeństwo obywatelskie: „Ograniczenia w przekraczaniu granic, ograniczenia zakupu (...) nieruchomości, podróżowania samolotem, uprawiania turystyki, wyjeżdżania na urlop i zatrzymywania się w hotelach wyższej klasy". W biuletynie ostrzega się również, że w przypadku „osób poważnie nadużywających zaufania władz" rząd będzie „instruował banki komercyjne, aby (...) ograniczały dostęp do kredytów, ubezpieczenia od utraty dochodów i innych podobnych usług"[27]. Lojalność popłaca, a nadużywanie zaufania władz jest naganne – i karalne. Rogier Creemers z Instytutu Van Vollenhovena na uniwersytecie w Lejdzie, specjalista w zakresie chińskiego prawa i zarządzania, tak to określił: „bękart systemu lojalnościowego"[28].

Nie mam zbyt wielu pocieszających słów dla „użytkowników" Sesame Credit, nie chciałabym jednak snuć samych czarnych wizji. Tu i ówdzie przebija się promyk nadziei. Nawet jeśli przyszłość wygląda nieciekawie, mamy przesłanki, by wierzyć, że zły trend się odwraca. Wiele osób z branży analitycznej od dawna wie o wykorzystywaniu osobistych danych dla zysku i od dawna się temu sprzeciwia. Dopóki jednak nie wybuchł wielki skandal z Cambridge Analytica, zagadnienia te nie były w centrum zainteresowania międzynarodowych mediów, bo rzadko trafiały na pierwsze strony gazet. Kiedy na początku 2018 roku wypłynęły brudy związane z kampanią wyborczą Trumpa, do opinii publicznej po raz pierwszy dotarło, że algorytmy po cichu gromadzą informacje o użytkownikach sieci – i stało się jasne, że bez nadzoru i regulacji takie praktyki mogą mieć dramatyczne konsekwencje.

Pewne regulacje już zaczynają wchodzić w życie. W Unii Europejskiej od niedawna obowiązuje nowe prawo, zwane RODO – rozporządzenie o ochronie danych osobowych – na mocy którego większość operacji przeprowadzanych obecnie przez brokerów danych jest sprzeczna z prawem. Przynajmniej w teorii nikt nie będzie już mógł przechowywać twoich danych bez wyraźnie określonego celu. Nie będzie mógł cię profilować bez twojej zgody. Nie będzie też mógł pozyskiwać twojej zgody na zbieranie danych w jednym celu, by potem potajemnie wykorzystywać je do czegoś zupełnie innego. Nie oznacza to jednak wcale, że te praktyki się skończą. Po pierwsze dlatego, że surfując po internecie, zazwyczaj nie tracimy czasu na czytanie warunków użytkowania programów i aplikacji, może się więc zdarzyć, że mimowolnie wyrazimy zgodę na gromadzenie danych wrażliwych. Po drugie, rozpoznawanie bezprawnej działalności i egzekwowanie nowego prawa nadal nie będzie łatwe w świecie, w którym większość analityki i transferu danych odbywa się w głębokim cieniu. Musimy po prostu poczekać – czas pokaże, jak się to wszystko ułoży.

Europejczycy znaleźli się w gronie szczęśliwców, ale również w Ameryce są podmioty zainteresowane wprowadzeniem podobnych regulacji. Federalna Komisja Handlu już w 2014 roku opublikowała raport potępiający szemrane praktyki brokerów danych i od tamtej pory aktywnie lobbuje za poszerzeniem wachlarza praw konsumentów. Firma Apple wbudowała w swoją przeglądarkę Safari „inteligentny system ochrony

przed śledzeniem", to samo zrobił Firefox. Facebook zrywa kontakty ze swoimi brokerami danych. Argentyna i Brazylia, Korea Południowa i wiele, wiele innych krajów forsuje regulacje podobne do naszego RODO. Europa dzisiaj jest w czołówce, ale ten pozytywny trend ma zasięg globalny.

Jeżeli dane są nowym złotem, to od dawna żyjemy na Dzikim Zachodzie. Jestem jednak optymistką i mam nadzieję, że wielu z nas będzie mogło wkrótce powiedzieć, że najgorsze już za nami.

Niemniej warto ciągle pamiętać, że nie ma nic za darmo. Legislacja nadrabia straty i społeczeństwo zaczyna wygrywać wojnę z chciwymi korporacjami, ale musimy być czujni – nie można dać się omamić złudą prywatności. Zanim użyjesz jakiegokolwiek algorytmu – zwłaszcza darmowego – zadaj sobie pytanie o ukryte motywy. Dlaczego ta aplikacja daje mi to wszystko za darmo? Co ten algorytm naprawdę robi? Czy ta wymiana mi się opłaci? A może jednak lepiej tego nie instalować?

Te lekcje możemy zastosować poza światem wirtualnym, ponieważ kalkulacje algorytmiczne przeniknęły praktycznie do wszystkich dziedzin życia społecznego. Żywiące się danymi algorytmy potrafią nie tylko przewidywać twoje zachowania w sklepie. Ich władza jest o wiele większa – mogą nawet pozbawić cię wolności.

Wymiar sprawiedliwości

To raczej norma, że w niedzielny letni wieczór na ulicach Brixton, gdzie zaczyna się nasza kolejna historia, można spotkać pozytywnie nastawione do życia grupki podchmielonych imprezowiczów. Brixton w południowym Londynie cieszy się reputacją dobrego miejsca na nocną zabawę; w tamten wieczór skończył się właśnie festiwal muzyczny i dzielnica roiła się od roześmianych ludzi – jedni wracali do domów, inni zamierzali jeszcze się bawić. Ale pół godziny przed północą nastrój się zmienił. W okolicznym osiedlu domów komunalnych wybuchła bójka. Kiedy policja nie zdołała opanować sytuacji, bijatyka szybko ogarnęła centrum Brixton i setki młodych ludzi włączyły się do zadymy.

Był sierpień 2011 roku. Poprzedniego wieczoru po drugiej stronie miasta pokojowa demonstracja po zastrzeleniu przez policję młodego mieszkańca Tottenhamu Marka Duggana przerodziła się w zamieszki. Teraz, drugą noc z rzędu, dzielnice Londynu zaczynały pogrążać się w chaosie – tym razem jednak atmosfera była inna: lokalna demonstracja w Tottenham dała początek masowemu łamaniu prawa i porządku.

Zamieszki rozkręcały się na dobre, kiedy Nicholas Robinson, 23-letni student elektrotechniki, wyszedł od swojej dziewczyny, z którą spędził weekend, i zwyczajowym skrótem przez Brixton ruszył do domu[1]. Znajome ulice zmieniły się praktycznie nie do poznania – przewrócone samochody, porozbijane okna, pożary i splądrowane sklepy[2]. Policjanci podejmowali rozpaczliwe wysiłki, żeby uspokoić sytuację, nie byli jednak w stanie zatrzymywać samochodów ani skuterów, które podjeżdżały pod wybite witryny sklepowe i odjeżdżały wyładowane ubraniami, butami, laptopami i telewizorami. Brixton całkowicie wymknęło się spod kontroli.

Kilka ulic za całkowicie splądrowanym sklepem elektrycznym Nicholas Robinson przechodził obok swojego supermarketu. Jak prawie wszystkie sklepy w okolicy, był całkowicie zrujnowany: okna zostały wybite, drzwi wyłamane, a regały w środku zdemolowane przez tłum. Niezatrzymywani

przez policję ludzie rozbiegali się na wszystkie strony z nowymi laptopami pod pachą. Nicholas poczuł pragnienie, wszedł więc do sklepu i wziął sobie zgrzewkę butelkowanej wody za trzy i pół funta. Właśnie wychodził spomiędzy regałów, kiedy do sklepu wkroczyła policja. Nicholas natychmiast sobie uzmysłowił, co zrobił, upuścił więc zgrzewkę i rzucił się do ucieczki[3].

W nocy z poniedziałku na wtorek cały kraj wstrzymał oddech w obawie przed kolejnymi zamieszkami. I rzeczywiście, wandale znowu wylegli na ulice[4]. Wśród nich znalazł się osiemnastoletni Richard Johnson. Zaintrygowany tym, co zobaczył w wiadomościach, chwycił zdecydowanie niepasującą do letniej pogody kominiarkę, wskoczył do samochodu i zajechał pod lokalne centrum handlowe. Zamaskowany, pobiegł do sklepu z grami komputerowymi, zagarnął ich całą torbę i wrócił do samochodu[5]. Miał jednak pecha – zaparkował auto dokładnie pod okiem kamery monitoringu. Tablica rejestracyjna pozwoliła bez trudu go namierzyć, a zarejestrowane materiały – szybko sformułować akt oskarżenia[6].

Richard Johnson i Nicholas Robinson zostali aresztowani za czyny popełnione podczas zamieszek. Obaj zostali oskarżeni o włamanie. Obaj stanęli przed sądem. Obaj przyznali się do winy. I na tym podobieństwo obu spraw się kończy.

Najpierw odbyła się rozprawa Nicholasa Robinsona, który stanął przed sądem w Camberwell niecały tydzień po incydencie. Pomimo znikomej wartości skradzionej przez niego zgrzewki wody, braku przeszłości kryminalnej oraz faktu, że był studentem wydziału dziennego, a także oświadczenia podsądnego, że wstydzi się swojego czynu, sędzia uznał, że jego zachowanie przyczyniło się do pogłębienia panującej tego dnia w Brixton atmosfery bezprawia, i – ku zgrozie zgromadzonej na sali rodziny Nicholasa – skazał go na sześć miesięcy więzienia[7].

Sprawa Johnsona weszła na wokandę w styczniu 2012 roku. Choć w dniu zamieszek Johnson miał wyraźny zamiar dokonania czynu bezprawnego – włożył przecież kominiarkę, żeby ukryć swoją tożsamość – i również przyczynił się do pogłębienia się atmosfery bezprawia, uniknął więzienia. Dostał wyrok w zawieszeniu i musiał odrobić dwieście godzin prac społecznych[8].

Zagadkowa niekonsekwencja

System wymiaru sprawiedliwości wie, że nie jest doskonały – i nawet nie próbuje doskonały się stać. Oszacowanie winy i określenie kary to nie matematyka, a sędzia nie jest maszyną liczącą. To właśnie dlatego wyrażenia „uzasadniona wątpliwość" czy „wystarczające powody" zajmują poczesne miejsce w języku prawniczym i właśnie dlatego odwołania stanowią tak ważną część przewodu sądowego; system sam przyznaje, że osiągnięcie absolutnej pewności nie jest możliwe.

Mimo to różnice w sposobie traktowania podsądnych – takich jak Nicholas Robinson i Richard Johnson – wydają się bardzo krzywdzące. Zbyt wiele jest indywidualnych czynników w każdej rozpatrywanej sprawie, żeby móc powiedzieć z całą pewnością, że wyrok jest „niesprawiedliwy", niemniej chcielibyśmy mieć nadzieję, że sędziowie będą – w granicach zdrowego rozsądku – konsekwentni w swoich decyzjach. Gdybyś miał siostrę bliźniaczkę lub gdybyś miała brata bliźniaka i gdybyście popełnili takie samo przestępstwo, liczylibyście chyba na to, że sąd wymierzy wam identyczną karę. Tylko czy rzeczywiście tak właśnie by się stało?

W latach siedemdziesiątych grupa amerykańskich badaczy spróbowała znaleźć odpowiedź na pewien wariant tego pytania[9]. Zamiast wykorzystać w eksperymencie przestępców bliźniaków (co byłoby w praktyce trudne i etycznie niepożądane), skonstruowano kilka hipotetycznych przypadków i zwrócono się do 47 sędziów ze stanu Wirginia z pytaniem, jak by postąpili w każdej z tych spraw. Możesz spróbować własnych sił na jednym z przykładów zaczerpniętych z tego badania. Jak ocenić poniższą sprawę?

> Osiemnastoletnia kobieta została zatrzymana za posiadanie marihuany, a następnie aresztowana razem ze swoim chłopakiem i siedmioma znajomymi. Liczne dowody wskazywały na to, że grupa wypaliła znaczne ilości narkotyku, znaleziono też dużo niewypalonej marihuany, choć nie w posiadaniu kobiety, przy niej bowiem nie znaleziono nic. Kobieta nie była wcześniej karana, pochodziła z porządnego domu z klasy średniej i była dobrą studentką; nie opierała się podczas zatrzymania, ale też nie wyraziła skruchy z powodu swojego postępowania.

Różnice w sentencjach były drastyczne. Dwudziestu dziewięciu sędziów orzekło, że podsądna jest niewinna, a 18 – że winna. Ośmiu sędziów z tej drugiej grupy zalecało objęcie kobiety nadzorem sądowym, czterech uznało, że najlepszym rozwiązaniem będzie grzywna, trzech orzekło i nadzór sądowy, i grzywnę, a trzech było zdania, że podsądna powinna pójść do więzienia.

Zatem los kobiety zależał wyłącznie od tego, przed kim miała szczęście (albo nieszczęście) stanąć w sali sądowej: na podstawie materiału dowodowego mogła się spodziewać każdej kary, od oczyszczenia z wszystkich zarzutów po karę pozbawienia wolności.

To z pewnością mocny cios dla każdego, kto wierzy w jednolitość orzeczeń. A jest jeszcze gorzej – sędziowie bowiem nie tylko nie zgadzają się ze sobą w ocenie spraw, lecz mają też skłonności do z a p r z e c z a n i a s a m y m s o b i e.

W bardziej współczesnym badaniu 81 brytyjskich sędziów spytano, czy zwolniliby za kaucją grupę fikcyjnych podsądnych[10]. Każda z tych hipotetycznych spraw miała własną, skonstruowaną przez badaczy historię; każdy oskarżony miał za sobą inną niż pozostali przeszłość kryminalną. Podobnie jak ich koledzy z Wirginii, sędziowie brytyjscy nie zdołali wydać zgodnego wyroku w żadnej z 41 spraw, które im przedstawiono[11]. Tym razem jednak w grupie 41 hipotetycznych spraw przedstawionych każdemu sędziemu znalazło się 7 przypadków, które pojawiły się dwukrotnie – za drugim razem zmieniono tylko personalia oskarżonych, żeby sędzia się nie zorientował, że ma do czynienia z dubletem. Zagranie poniżej pasa, ale warto było: przy drugiej identycznej sprawie większość sędziów nie zdołała wydać takiego samego orzeczenia jak przy pierwszej. Zdumiewające jest to, że niektórzy wydali orzeczenia tak diametralnie różne, że równie dobrze mogliby – najzupełniej dosłownie – rzucać kostką do gry[12].

Wiele innych eksperymentów przywiodło badaczy do jednego wniosku: kiedy sędziowie będą mieli wolną rękę w ocenie sprawy, będą się pojawiać ogromne niezgodności. Pozostawienie sędziom swobody decyzji oznacza akceptację przypadkowości na salach sądowych.

Istnieje oczywiście prosty sposób na zagwarantowanie jednolitości orzecznictwa – wystarczy zlikwidować dyskrecjonalność sędziowską. Gdyby każdy oskarżony o to samo przestępstwo był sądzony dokładnie tak samo, mielibyśmy zapewnioną precyzję – przynajmniej w kwestii orzekania. Niektóre kraje już poszły tą drogą. Preskryptywne systemy orzecznictwa wprowadzono w Stanach Zjednoczonych na poziomie federalnym oraz w niektórych regionach Australii[13]. Tego rodzaju precyzja ma jednak swoją cenę. Za wzrost jednolitości w orzecznictwie trzeba zapłacić utratą innego rodzaju sprawiedliwości.

Zilustruję to na przykładzie. Wyobraź sobie dwóch podsądnych, którym zarzucono kradzież w supermarkecie. Jeden z nich to stosunkowo dobrze ustawiony w kryminalnym światku wykonawca robót zleconych; drugi to obecnie bezrobotny i przepełniony wyrzutami sumienia ojciec rodziny, który z trudem wiąże koniec z końcem i kradnie, żeby wykarmić dzieci. Jeśli odbierzemy sędziom możliwość brania pod rozwagę tego rodzaju okoliczności łagodzących i wprowadzimy ścisłe zalecenia, by wszystkie osoby oskarżone o takie samo przestępstwo traktować identycznie, możemy doprowadzić do tego, że niektórym wymierzymy karę zbyt surową, przekreślając w pewnych przypadkach szansę na resocjalizację.

Trudny dylemat. Bez względu na to, jaki system ustanowimy dla swoich sędziów, będzie trzeba wypracować kruchą równowagę między indywidualizacją wymiaru kary a jednolitością orzeczeń. Większość państw próbowała rozwiązać ten problem, wstępując na drogę złotego środka między preskryptywną skrajnością amerykańskiego prawa federalnego a sądownictwem opartym niemal wyłącznie na swobodzie sędziów, jak na przykład w Szkocji[14]. W całym świecie zachodnim wskazówki orzecznicze sugerują wymierzanie albo najwyższego wymiaru kary (jak w Irlandii), albo najniższego (jak w Kanadzie), lub wskazują widełki między najwyższym i najniższym wymiarem, pozostawiając sędziom swobodę w tak wyznaczonym zakresie[15].

Nie ma systemów doskonałych. Ale właśnie w takiej bardzo złożonej, chaotycznej sytuacji algorytm ma szansę pokazać się z najlepszej strony. Bo tak się wspaniale składa, że gdy zaprzęgniemy do pracy komputer, możemy zagwarantować w systemie zarówno jednolitość orzeczeń, jak i indywidualizację wymiaru kary. Nie musimy już wybierać mniejszego zła.

Sprawiedliwe równanie

Algorytmy nie mogą orzekać o winie. Nie mogą ważyć argumentów obrony i oskarżenia, analizować dowodów ani decydować o tym, czy podsądny odczuwa prawdziwą skruchę. Nie spodziewaj się zatem, że w najbliższej przyszłości zastąpią nam sędziów. Algorytm może za to – choć brzmi to niewiarygodnie – wykorzystać informacje o oskarżonym i obliczyć prawdopodobieństwo powtórnego popełnienia przez niego przestępstwa. A ponieważ wielu sędziów przed wydaniem orzeczenia rozważa także to, czy oskarżony może wrócić na ścieżkę bezprawia, jest to funkcja całkiem użyteczna.

Z danych i algorytmów korzysta się w sądownictwie już niemal od stulecia – po raz pierwszy użyto ich w Ameryce w latach dwudziestych ubiegłego wieku. W owym czasie w wymiarze sprawiedliwości Stanów Zjednoczonych standardowo zasądzano najwyższą przewidzianą karę, a po upływie jej określonej części oskarżony nabywał prawo do warunkowego zwolnienia. Dzięki tej praktyce dziesiątki tysięcy więźniów wcześniej wyszło na wolność. Jedni zostali skutecznie zresocjalizowani, inni nie. Wszyscy jednak stali się idealnym punktem wyjścia do przeprowadzenia naturalnego eksperymentu: czy można z góry przewidzieć, że skazany nie dotrzyma warunków przedterminowego zwolnienia?

I tu na scenę wkracza Ernest W. Burgess, zakochany w prognozach kanadyjski socjolog z University of Chicago. Burgess był wielkim zwolennikiem pomiarów zjawisk społecznych: przez całe życie zawodowe usiłował przewidywać wszystko – od skutków przejścia na emeryturę po trwałość małżeństw – i w 1928 roku stworzył pierwsze narzędzie pozwalające na skuteczne wyliczenie ryzyka popełnienia czynu zabronionego.

Analizując rozmaite informacje o trzech tysiącach osadzonych z trzech więzień stanu Illinois, Burgess zidentyfikował 21 czynników, które uznał za „potencjalnie znaczące" dla określenia prawdopodobieństwa naruszenia przez więźnia warunków przedterminowego zwolnienia. Wśród nich znalazły się między innymi: rodzaj przestępstwa, liczba miesięcy spędzonych w więzieniu oraz typ społeczny osadzonego, który – z taktem

typowym dla badacza zjawisk społecznych z początku XX wieku – Burgess wyrażał takimi pojęciami, jak „włóczęga", „pijak", „obibok", „chłopak ze wsi" i „imigrant"[16].

Burgess dał każdemu więźniowi od zera do jednego punktu w każdej z 21 kategorii i stwierdził, że dla wysokich wyników (łącznie między 16 a 21 punktów) prawdopodobieństwo recydywy było najmniejsze, natomiast więźniowie, którzy uzyskali niski wynik łączny (4 punkty lub mniej), byli jego zdaniem bardziej skłonni do naruszenia warunków zwolnienia.

Kiedy wszyscy osadzeni z jego bazy danych zostali w końcu przedterminowo zwolnieni, Burgess mógł się przekonać, jak trafne były jego prognozy. Jak na tak nieskomplikowaną analizę, wyniki były zdumiewająco dobre. Dziewięćdziesiąt procent więźniów z jego grupy niskiego ryzyka dotrwało do końca zwolnienia z czystym kontem, podczas gdy dwie trzecie osadzonych z jego grupy wysokiego ryzyka rzeczywiście naruszyło warunki[17]. Okazało się, że nawet prymitywne modele statystyczne potrafią przewidywać lepiej niż eksperci.

Nie brakowało jednak krytyków podejścia Burgessa. Sceptycy zastanawiali się, w jakim stopniu czynniki, które pozwoliły trafnie przewidzieć brak recydywy w określonym miejscu i czasie, mogą mieć zastosowanie w innych okolicznościach. (Coś było na rzeczy: nie wiem, czy kategoria „chłopak ze wsi" przydałaby się nam dzisiaj w prognozowaniu recydywy w grupie współczesnych przestępców ze starych dzielnic miasta). Inni badacze krytykowali Burgessa za to, że po prostu wykorzystał dostępne informacje, nie sprawdziwszy najpierw, jak dalece są istotne[18]. Pytano go również o sposób przydzielania punktów: jego metoda polegała w zasadzie na zapisywaniu własnych (czyli subiektywnych) opinii w postaci formuł matematycznych. Mimo tych głosów krytycznych możliwości prognostyczne narzędzia Burgessa były tak imponujące, że w 1935 roku jego metodą zaczęły się posiłkować więzienne komisje do spraw warunkowych zwolnień w stanie Illinois[19]. A na przełomie tysiącleci matematyczni spadkobiercy metody Burgessa znaleźli zatrudnienie na całym świecie[20].

Przenieśmy się szybko w czasie do dnia dzisiejszego – najnowocześniejsze algorytmy oceny ryzyka używane w salach sądowych

są znacznie bardziej wyrafinowane od prymitywnych narzędzi zaprojektowanych przez Burgessa. Nie tylko asystują przy wydawaniu decyzji o warunkowych zwolnieniach, lecz także pomagają przydzielać więźniom programy interwencyjne i rozstrzygać, kogo zwolnić z aresztu za kaucją, a całkiem od niedawna wspierają również sędziów w orzecznictwie. Fundamentalne zasady pozostały te same: komputer dostaje podstawowe dane – wiek oskarżonego, jego przeszłość kryminalną, klasyfikację czynu zabronionego i tak dalej – i wylicza swoją numeryczną prognozę: w jakim stopniu bezpieczne byłoby wypuszczenie takiej osoby na wolność.

Na jakiej zasadzie działają te współczesne algorytmy? Na ogół najlepiej sprawdzające się maszyny wykorzystują technikę zwaną lasem losowym, której podstawą jest genialnie prosta idea: skromne drzewo decyzyjne.

Pytanie do publiczności

Możesz pamiętać drzewa decyzyjne ze szkoły. To popularny wśród nauczycieli matematyki sposób strukturyzowania obserwacji przy takich eksperymentach, jak rzucanie monetą czy kostką do gry. Gotowe drzewo decyzyjne da się też wykorzystać w charakterze schematu blokowego: zakładamy warunki początkowe i krok po kroku decydujemy, co robić albo – tak jak w naszym wypadku – co się teraz wydarzy.

Wyobraź sobie, że chcesz zdecydować, czy zwolnić za kaucją określoną osobę. Podobnie jak w przypadku warunkowego zwolnienia, decyzja opiera się na prostej kalkulacji. Wina nie ma tu znaczenia. Masz po prostu dać odpowiedź na pytanie, czy oskarżony naruszy warunki zwolnienia, jeżeli zostanie wypuszczony po wpłaceniu poręczenia majątkowego.

Pomocne będą informacje dotyczące kilkunastu innych oskarżonych – część z nich po zwolnieniu za kaucją uciekła i nie stawiła się na rozprawie albo do czasu rozprawy ponownie złamała prawo, a część nie. Załóżmy teraz, że korzystając z tych danych, narysujesz odręcznie proste drzewo decyzyjne, takie jak to poniżej. Dzięki niemu będzie można z dużym prawdopodobieństwem przewidzieć, jak się zachowa dany oskarżony. Przesuwaj się po polach odpowiadających jego charakterystyce, aż

dojdziesz do pola z prognozą. Dopóki każdy będzie szedł tą samą ścieżką co jego poprzednik, prognoza będzie trafna.

Tutaj jednak takie drzewa decyzyjne, jakie rysowaliśmy w szkole, zaczynają się przewracać. Bo oczywiście n i e k a ż d y pójdzie tą samą ścieżką co jego poprzednicy. Pojedyncze drzewo będzie generować wiele złych prognoz. I nie tylko dlatego że zaczęliśmy od prostego przykładu. Nawet gdybyśmy mieli potężną bazę danych poprzednich podsądnych i niewiarygodnie skomplikowany schemat blokowy, metoda ta ciągle będzie prawdopodobnie tylko trochę lepsza od zwykłego rzucania kostką do gry.

Jeżeli jednak skonstruujesz więcej drzew, wszystko się zmieni. Zamiast wykorzystywać wszystkie dane od razu – podziel je i rządź! Zbuduj najpierw tysiące mniejszych drzewek z przypadkowo wydzielonych podzbiorów danych – uzyskasz tak zwany z e s p ó ł s t a t y s t y c z n y. I kiedy dostaniesz do oceny nowego oskarżonego, po prostu poprosisz każde drzewo z osobna o zagłosowanie w ankiecie, czy wypuszczenie go z aresztu za poręczeniem majątkowym to dobry czy zły pomysł. Nie wszystkie drzewa będą tego samego zdania i w pojedynkę będą z pewnością generować złe prognozy, ale kiedy weźmiesz średnią ze

wszystkich odpowiedzi, jakie dostaniesz, zdecydowanie poprawisz trafność swoich przewidywań.

Przypomina to trochę jedno z kół ratunkowych w *Milionerach*: pytanie do publiczności. Sala wypełniona wieloma obcymi ludźmi będzie miała rację częściej niż twoja najbardziej oczytana przyjaciółka. (Koło ratunkowe „pytanie do publiczności" ma 91-procentową skuteczność w porównaniu z zaledwie 65 procentami „telefonu do przyjaciela"[21]). Błędy popełnione przez członków grupy będą się wzajemnie znosić, dzięki czemu grupa będzie mądrzejsza od jednostki.

Tak samo rzecz się ma z dużą grupą drzew decyzyjnych, które razem tworzą właśnie las losowy. Ponieważ prognozy algorytmu bazują na wzorcach, których ten się uczy, kiedy analizuje dane, las losowy określa się mianem algorytmu uczenia maszynowego – które z kolei podpada pod szerokie pojęcie sztucznej inteligencji. (Pojęcie uczenia maszynowego pojawiło się już w rozdziale *Władza*, ale w dalszej części książki poznamy jeszcze wiele algorytmów należących do tej kategorii, tymczasem jednak warto odnotować, że to naprawdę niesamowite, gdy zdamy sobie sprawę, iż taki inteligentny algorytm to właściwie grupa znanych ze szkoły schematów blokowych posklejanych klajstrem manipulacji matematycznych). Lasy losowe potwierdziły swoją niewiarygodną użyteczność na bardzo wielu polach w świecie rzeczywistym. Używa ich Netflix, który na podstawie twoich dotychczasowych upodobań przewiduje, co chciałbyś obejrzeć[22]; używa ich Airbnb do wykrywania fałszywych kont[23]; używa ich służba zdrowia do diagnozowania chorób (więcej na ten temat w następnym rozdziale).

Trzeba przyznać, że kiedy używa się ich do oceny oskarżonych, w dwóch rzeczach biją na głowę człowieka w sędziowskiej todze. Po pierwsze, algorytm udzieli zawsze identycznej odpowiedzi, kiedy zostanie mu zaprezentowany identyczny zestaw okoliczności. Jednolitość orzeczeń gwarantowana, i to bynajmniej nie kosztem indywidualizacji wymiaru kary. I jest jeszcze druga niemała korzyść: algorytm jest lepszym prognostą niż człowiek.

Człowiek kontra maszyna

W 2017 roku grupa badaczy postanowiła sprawdzić, jak wypadną przewidywania maszyny w zestawieniu z opiniami sędziów[24].

Eksperymentatorzy dostali materiał w postaci kartotek wszystkich osób aresztowanych w Nowym Jorku w ciągu pięciu lat: od 2008 do 2013 roku. W tym czasie przesłuchanie w sprawie kaucji miało 750 tysięcy osób, a to całkiem spory materiał do przetestowania algorytmu w bezpośrednim starciu z żywym sędzią.

Wymiar sprawiedliwości Nowego Jorku nie korzystał w tych rozprawach z żadnego algorytmu, ale badacze stworzyli szereg drzew decyzji do oceny dokładności, z jaką można było wówczas określić ryzyko złamania warunków zwolnienia. Algorytm dostawał zwyczajowe dane: kartotekę, popełnione przestępstwo i tak dalej, a zwracał prognozę zachowania się podsądnego po wyjściu z aresztu.

Twarde dane mówiły o tym, że z aresztu zwolniono za kaucją 408 283 oskarżonych. Każda z tych osób mogła uciec i nie stawić się na właściwej rozprawie lub popełnić na wolności inne przestępstwo; oznacza to, że korzystając z kartotek sądowych, możemy ustalić, jak dokładne były prognozy algorytmu i jak trafne decyzje podjęte przez sędziów. Wiemy, ile osób nie stawiło się później w sądzie na właściwej rozprawie (15,2 procent), a ile jeszcze przed właściwą rozprawą zostało ponownie aresztowanych za popełnienie innego przestępstwa (25,8 procent).

Niedobre z punktu widzenia statystyki było to, że żaden oskarżony, którego sędzia zakwalifikował do grupy wysokiego ryzyka, nie został zwolniony z aresztu za poręczeniem majątkowym – i w tych przypadkach badacze nie mogli określić, czy sędzia się pomylił, czy też miał rację. To trochę komplikuje sprawy. To znaczy, że nie ma możliwości przedstawienia twardych danych liczbowych pokazujących ogólną trafność decyzji wszystkich sędziów. A bez wiedzy o tym, jak oskarżeni zachowywaliby się na wolności, nie będziemy w stanie określić ogólnej trafności algorytmu. Trzeba zatem wypracować jakiś sposób na przewidywanie, co zatrzymani w areszcie podsądni zrobiliby na wolności, gdyby zostali zwolnieni z aresztu[25], i obliczyć wynik pojedynku człowieka z maszyną trochę okrężną drogą.

Mimo to jedno jest pewne: orzeczenia sędziów nie były zgodne z obliczeniami maszyny. Badacze pokazali, że wielu podsądnych, których algorytm uznał za prawdziwych kryminalistów, zostało potraktowanych przez sędziów jak osoby niskiego ryzyka. Prawie połowa oskarżonych zakwalifikowanych przez maszynę do grupy największego ryzyka została przez sędziów zwolniona za kaucją.

No i kto miał rację? Dane wykazały, że wypuszczenie na wolność osób, co do których algorytm miał spore zastrzeżenia, rzeczywiście było decyzją ryzykowną. Ponad 56 procent z nich nie stawiło się na właściwej rozprawie w sądzie, a 62,7 procent popełniło przed terminem rozprawy kolejne przestępstwo – w tym najgorsze zbrodnie: gwałt i zabójstwo. Algorytm te zagrożenia wyraźnie dostrzegał.

Badacze orzekli, że bez względu na sposób wykorzystania algorytmu jego prognozy będą znacznie trafniejsze od przewidywań sędziów. I mają konkretne liczby na poparcie swoich argumentów. Jeżeli chcesz zostawić w areszcie mniej osób oczekujących na rozprawę, algorytm pójdzie ci na rękę – pośle za kratki o 41,8 procent mniej oskarżonych niż człowiek, utrzymując przy tym przestępczość na niezmienionym poziomie. A jeżeli bylibyśmy zadowoleni z obecnego odsetka zwolnionych za kaucją, algorytm też ci przyjdzie w sukurs: dzięki większej trafności przy wyborze osób, które wyjdą na wolność, algorytm pomoże ci ograniczyć liczbę uciekinierów o 24,7 procent.

Te korzyści to już nie teoria. Rhode Island, gdzie sądy używają takich algorytmów już od ośmiu lat, może się poszczycić 17-procentowym spadkiem liczby osadzonych i 6-procentowym spadkiem przypadków recydywy. To przekłada się w praktyce na setki podsądnych niskiego ryzyka, którzy nie siedzą niepotrzebnie w więzieniu w oczekiwaniu na rozprawę, oraz na setki przestępstw i zbrodni, do których nigdy nie doszło. A ponieważ roczny koszt utrzymania jednego więźnia w Wielkiej Brytanii przekracza 30 tysięcy funtów[26] (w Stanach Zjednoczonych spędzenie roku w więzieniu o zaostrzonym rygorze może kosztować tyle co rok studiów na Harvardzie[27]) – oznacza to również ogromne oszczędności dla podatników. Tak więc wszyscy są zadowoleni, w tym układzie nie ma przegranych.

Ale czy na pewno?

Gdzie jest Darth Vader?

Oczywiście żaden algorytm nie zdoła b e z b ł ę d n i e przewidzieć, co dany człowiek zrobi w przyszłości. Ludzie są zbyt chaotyczni, irracjonalni i impulsywni, żeby jakikolwiek prognosta mógł być kiedykolwiek p e w n y, co się wydarzy za chwilę. Może dawać lepsze prognozy od innych, nadal jednak będzie popełniał błędy. Powstaje więc pytanie: co się dzieje z tymi wszystkimi ludźmi, którym źle obliczono ryzyko popełnienia przestępstwa?

Są dwa rodzaje błędów, jakie może popełnić algorytm. Richard Berk, profesor kryminologii i statystyki na University of Pennsylvania i pionier w dziedzinie oceny ryzyka recydywy, potrafi to wytłumaczyć w bardzo oryginalny sposób:

„Na świecie są dobrzy goście i źli goście. I dlatego twój algorytm w każdym przypadku zadaje sobie to samo pytanie: «Darth Vader czy Luke Skywalker?»".

Zwolnienie z aresztu Dartha Vadera to pierwszy rodzaj błędu, znany jako „w y n i k f a ł s z y w i e n e g a t y w n y". Algorytm popełnia go za każdym razem, kiedy nie rozpozna ryzyka związanego z wypuszczeniem na wolność groźnego przestępcy.

Z drugiej strony uwięzienie Luke'a Skywalkera też nie będzie dobre; to błąd zwany „w y n i k i e m f a ł s z y w i e p o z y t y w n y m". Algorytm popełnia go wówczas, gdy niepoprawnie uznaje podsądnego za osobę wysokiego ryzyka.

Oba te błędy nie są związane wyłącznie z oceną ryzyka recydywy; będą raz za razem wypływać na kartach tej książki. Każdy algorytm klasyfikujący może zostać obwiniony o ich popełnienie.

Berk twierdzi, że jego algorytmy mogą przewidywać z dokładnością dochodzącą do 75 procent, czy dany oskarżony dopuści się zabójstwa – co stawia je w rzędzie najdokładniejszych na świecie[28]. Jeżeli weźmiemy pod uwagę, jak chętnie mówimy o wolnej woli człowieka, jest to dokładność naprawdę imponująca. Lecz nawet przy tak wielkiej dokładności znajdzie się niejeden Luke Skywalker, który nie zostanie zwolniony z aresztu za kaucją, bo za bardzo przypomina Dartha Vadera.

Konsekwencje niewłaściwego klasyfikowania podsądnych stają się jeszcze poważniejsze, kiedy algorytmów używa się nie tylko przy podejmowaniu decyzji podczas przesłuchania w sprawie zwolnienia za kaucją, lecz także przy formułowaniu orzeczeń. To już nasza rzeczywistość: od całkiem niedawna w niektórych amerykańskich stanach sędziowie mogą ustalać długość kary pozbawienia wolności skazanych przestępców, posiłkując się obliczonym przez algorytm stopniem ryzyka recydywy. Dopuszczenie takiej praktyki rozpętało żywiołową debatę i nie ma się czemu dziwić: co innego obliczać, czy można podsądnego zwolnić z aresztu przed terminem rozprawy, a co innego obliczać, na jak długo wsadzić go za kratki.

Problem po części w tym, że przy ustalaniu długości odsiadki należy rozważyć znacznie więcej niż samo ryzyko powrotu na drogę przestępstwa – a tylko to może obliczyć algorytm. Sędzia musi uwzględnić zagrożenie, jakie oskarżony może stanowić dla innych, efekt odstraszający wydanego w jego sprawie wyroku, kwestię zadośćuczynienia wobec ofiary oraz szanse na resocjalizację sprawcy. Trudno to wszystko harmonijnie pogodzić, nic więc dziwnego, że podnoszą się głosy sprzeciwu, kiedy algorytm zaczyna mieć za duży wpływ na brzmienie wyroku. Nie można się też dziwić, gdy przeciwnicy tego rozwiązania przytaczają bardzo niepokojący casus Paula Zilly'ego[29].

Zilly został skazany za kradzież kosiarki. W lutym 2013 roku stanął przed sędzią Bablerem w hrabstwie Baron w stanie Wisconsin ze świadomością, że jego obrońcy zawarli porozumienie z oskarżycielem i w zamian za złagodzenie zarzutów Zilly przyzna się do winy. Obie strony zgadzały się co do tego, że długa odsiadka nie byłaby w tej sprawie najlepszym rozwiązaniem. Zilly stawił się więc w sądzie przekonany, że sędzia po prostu przyklepie umowę, jaką zawarł jego adwokat z prokuraturą.

Tak się jednak składa, że sędziowie w Wisconsin używali już wówczas własnościowego algorytmu oceny ryzyka recydywy o nazwie COMPAS. Podobnie jak to było z narzędziem budżetowym stanu Idaho w rozdziale *Władza*, zasada działania COMPAS-u stanowi tajemnicę handlową – z tą różnicą, że kod sędziowskiego algorytmu z Wisconsin do tej pory nie został ujawniony opinii publicznej. Wiemy tyle, że obliczenia są wykonywane na podstawie odpowiedzi oskarżonego na pytania

z kwestionariusza. Takich jak: „Człowiek głodny ma prawo ukraść, tak czy nie?" albo „Jeżeli mieszkałeś/mieszkałaś z rodzicami i doszło do ich rozwodu, ile miałeś/miałaś wtedy lat?"[30]. Algorytm został zaprojektowany wyłącznie w jednym celu – do przewidywania ryzyka ponownego popełnienia przestępstwa przez oskarżonego w ciągu dwóch lat, przy czym osiągnął wskaźnik trafności na poziomie 70 procent[31]. To oznacza mniej więcej tyle, że w co trzecim przypadku maszyna podejmuje błędne decyzje, ale mimo to sędziowie z Wisconsin posiłkowali się nią przy wydawaniu wyroków.

Wynik Zilly'ego nie był dobry. Algorytm wyliczył mu wysokie ryzyko popełnienia w przyszłości przestępstwa z użyciem przemocy oraz średnie ryzyko recydywy. „Patrzę na ocenę ryzyka – powiedział sędzia Babler na sali sądowej – i trudno o gorszy wynik".

Sędzia bardziej zaufał maszynie niż porozumieniu wypracowanemu przez obrońcę i oskarżyciela – odrzucił wniosek stron i podwoił pierwotną karę: zamiast roku w więzieniu okręgowym Zilly dostał dwa lata odsiadki w więzieniu stanowym o zaostrzonym rygorze.

Nigdy się nie dowiemy, czy Zilly zasłużył na tak wysoki wynik w ocenie ryzyka, niemniej dokładność na poziomie 70 procent wydaje się zdecydowanie za niska, żeby usprawiedliwiać stawianie wskazań algorytmu ponad innymi czynnikami.

Sprawa Zilly'ego była bardzo głośna, nie był to jednak bynajmniej odosobniony przypadek. W 2003 roku dziewiętnastoletni Christopher Drew Brooks odbył konsensualny stosunek płciowy z czternastolatką. Sąd w stanie Wirginia uznał go jednak za winnego kontaktu seksualnego z osobą nieletnią. Wskazówki orzecznicze przewidywały za to od 7 do 16 miesięcy pozbawienia wolności, ale po uwzględnieniu oceny ryzyka (nie przez algorytm COMPAS) sugerowany najwyższy wymiar kary został podwyższony do 24 miesięcy. Sędzia uwzględnił to w orzeczeniu i skazał Brooksa na 18 miesięcy więzienia[32].

Tutaj użyty przez sąd algorytm przy obliczaniu ryzyka recydywy uwzględniał również wiek podsądnego. Oskarżenie o przestępstwo seksualne w tak młodym wieku przemawiało przeciw Brooksowi, mimo że był zbliżony wiekiem do ofiary. I faktycznie – gdyby sprawca miał 36 lat

(a zatem gdyby był o 22 lata starszy od dziewczynki), algorytm nawet by nie sugerował kary pozbawienia wolności[33].

Nie są to pierwsze przypadki, kiedy człowiek przedkłada wyniki obliczeń komputera nad własny osąd – i z pewnością nie będą to przypadki ostatnie. I znów musimy zadać pytanie: co możemy z tym zrobić? Głos w sprawie zabrał Sąd Najwyższy stanu Wisconsin. Odnosząc się do zagrożeń związanych z przykładaniem przez sędziów zbyt wielkiej wagi do wskazań komputerów, stwierdził, co następuje: „Spodziewamy się, że sądy okręgowe będą zachowywały powściągliwość względem wyników oceny ryzyka algorytmu COMPAS w odniesieniu do każdego podsądnego”[34]. Richard Berk mówi jednak, że to nadmiar optymizmu: „Sędziowie bardzo się boją popełnić błąd – dotyczy to szczególnie tych wyznaczonych przez obywateli. A dzięki algorytmom nie muszą się tak bardzo wysilać, unikając przy tym osobistej odpowiedzialności”[35].

I tu powstaje kolejny problem. Kiedy algorytm zaklasyfikuje oskarżonego do grupy wysokiego ryzyka i w rezultacie sędzia pozbawi tego człowieka wolności, nie będziemy mogli z całą pewnością stwierdzić, czy algorytm trafnie przewidział przyszłość podsądnego. Weźmy przykład Zilly’ego. Być może rzeczywiście popełniłby później przestępstwo z użyciem przemocy. A być może nie. Kto wie, czy przez to, że został uznany za recydywistę wysokiego ryzyka i wysłany do więzienia stanowego, nie wszedł na ścieżkę inną od tej, na której już był, kiedy jego obrońca zawarł porozumienie z oskarżycielem. A ponieważ nie dysponujemy żadną metodą weryfikacji prognozy algorytmu, nie mamy też żadnej możliwości stwierdzenia, czy sędzia postąpił słusznie, dając pierwszeństwo ocenie ryzyka – i dlatego nigdy się nie dowiemy, czy Zilly był rzeczywiście Darthem Vaderem, czy może jednak Lukiem Skywalkerem.

To problem, który nie ma łatwego rozwiązania. Bo jak przekonać sędziego, żeby patrząc na wyniki algorytmu, słuchał głosu rozsądku? Poza tym – nawet gdyby się to udało – jest jeszcze jeden kłopot z prognozowaniem ryzyka recydywy. I to zdecydowanie bardziej kontrowersyjny.

Uprzedzona maszyna

W 2016 roku niezależny internetowy portal informacyjny ProPublica, który pierwszy opisał historię Zilly'ego, przyjrzał się szczegółowo COMPAS-owi i za pomocą inżynierii odwrotnej zbadał jego prognozy dotyczące siedmiu tysięcy oskarżonych rzeczywiście sądzonych na Florydzie w latach 2013–2014[36]. Badacze chcieli się przekonać, jak dokładne były oceny ryzyka recydywy generowane przez algorytm, sprawdzili więc, ile osób rzeczywiście popełniło kolejne przestępstwo. Drugim celem badania było sprawdzenie, czy oceny były różne, czy takie same dla oskarżonych o czarnym i białym kolorze skóry.

Algorytm nie bierze pod uwagę rasy podsądnych, dziennikarze ProPubliki odkryli jednak, że w obliczeniach nie wszyscy byli traktowani równo. Mimo że prawdopodobieństwo popełnienia błędu przez maszynę było takie samo dla białych i czarnych oskarżonych, algorytm mylił się inaczej wobec każdej z obu tych grup.

Jeżeli podsądny należał do grona osób, które po pierwszym zatrzymaniu nie weszły ponownie w konflikt z prawem – innymi słowy, jeśli był Lukiem Skywalkerem – prawdopodobieństwo fałszywego zaliczenia go przez algorytm do grupy wysokiego ryzyka było dwukrotnie większe, jeśli był czarnoskóry. Generowane przez algorytm wyniki fałszywie pozytywne były nieproporcjonalnie czarne. I na odwrót – w grupie oskarżonych, którzy faktycznie dopuścili się kolejnego przestępstwa w ciągu dwóch lat od zatrzymania – w grupie Darth Vaderów – prawdopodobieństwo postawienia fałszywej prognozy niskiego ryzyka było dwukrotnie większe dla podsądnych o białym kolorze skóry niż dla czarnoskórych. Generowane przez algorytm wyniki fałszywie negatywne były nieproporcjonalnie białe.

Trudno się dziwić, że analiza ProPubliki wywołała wielki skandal w całych Stanach Zjednoczonych i poza granicami kraju. Prasa wydrukowała setki artykułów wyrażających zdecydowaną dezaprobatę dla wykorzystywania bezdusznych kalkulatorów do wymierzania sprawiedliwości ludziom; ich autorzy miażdżyli krytyką praktykę stosowania niedoskonałych, rasowo uprzedzonych algorytmów w podejmowaniu decyzji, które mogą przekreślić przyszłość człowieka. Z wieloma z tych argumentów należy się zgodzić – każdy oskarżony

zasługuje na uczciwy proces i równe traktowanie niezależnie od tego, kto jest sędzią w jego sprawie. Pod tym względem algorytm w badaniu ProPubliki wypadł bardzo źle.

Powstrzymajmy się jednak na chwilę i nie wymierzajmy jeszcze sprawiedliwości wszystkim „niedoskonałym algorytmom". Zanim zabronimy stosowania jakichkolwiek algorytmów w sądownictwie, zadajmy sobie pytanie: jak ma wyglądać nieuprzedzony algorytm?

Bez wątpienia chcielibyśmy, żeby obliczał trafnie ryzyko dla białych i czarnych podsądnych. Rozsądne wydaje się również żądanie, żeby jego kategoria „wysokiego ryzyka" była identyczna dla wszystkich; algorytm powinien z jednakową trafnością wyłapywać potencjalnych recydywistów, bez względu na ich przynależność do takiej czy innej grupy, nie tylko etnicznej. No i jeszcze – na co zwrócili uwagę dziennikarze ProPubliki – algorytm powinien popełniać takie same błędy z taką samą częstotliwością wobec przedstawicieli wszystkich ras.

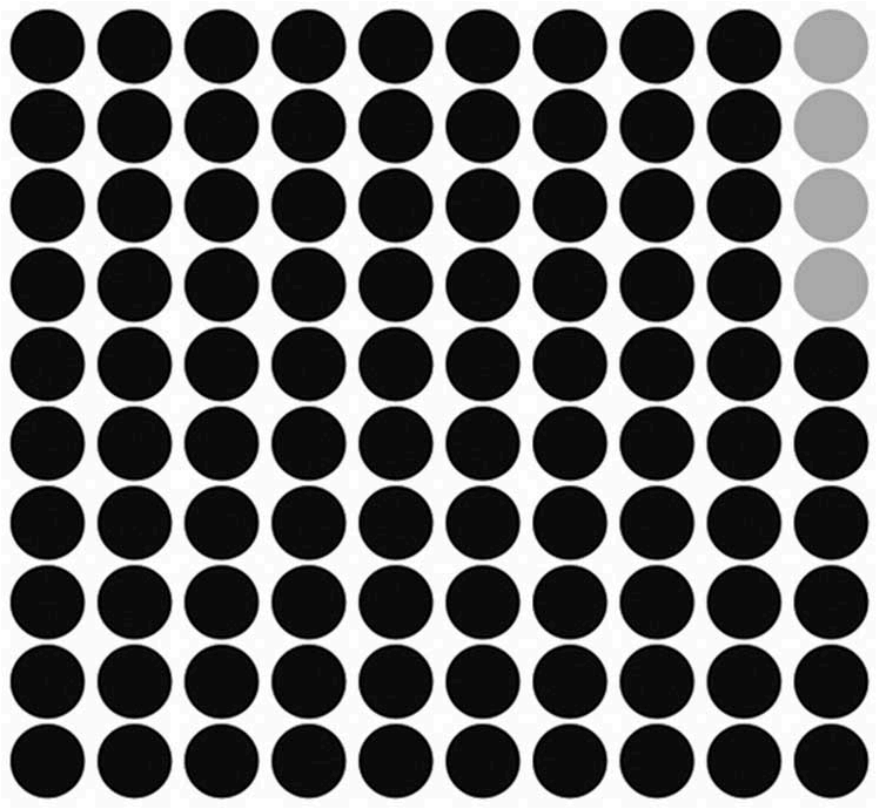
Żaden z tych czterech postulatów nie sprawia wrażenia nieziszczalnego marzenia. Mamy tu jednak pewien problem, bo niestety jest tak, że pewne rodzaje równości są matematycznie niekompatybilne z innymi jej rodzajami.

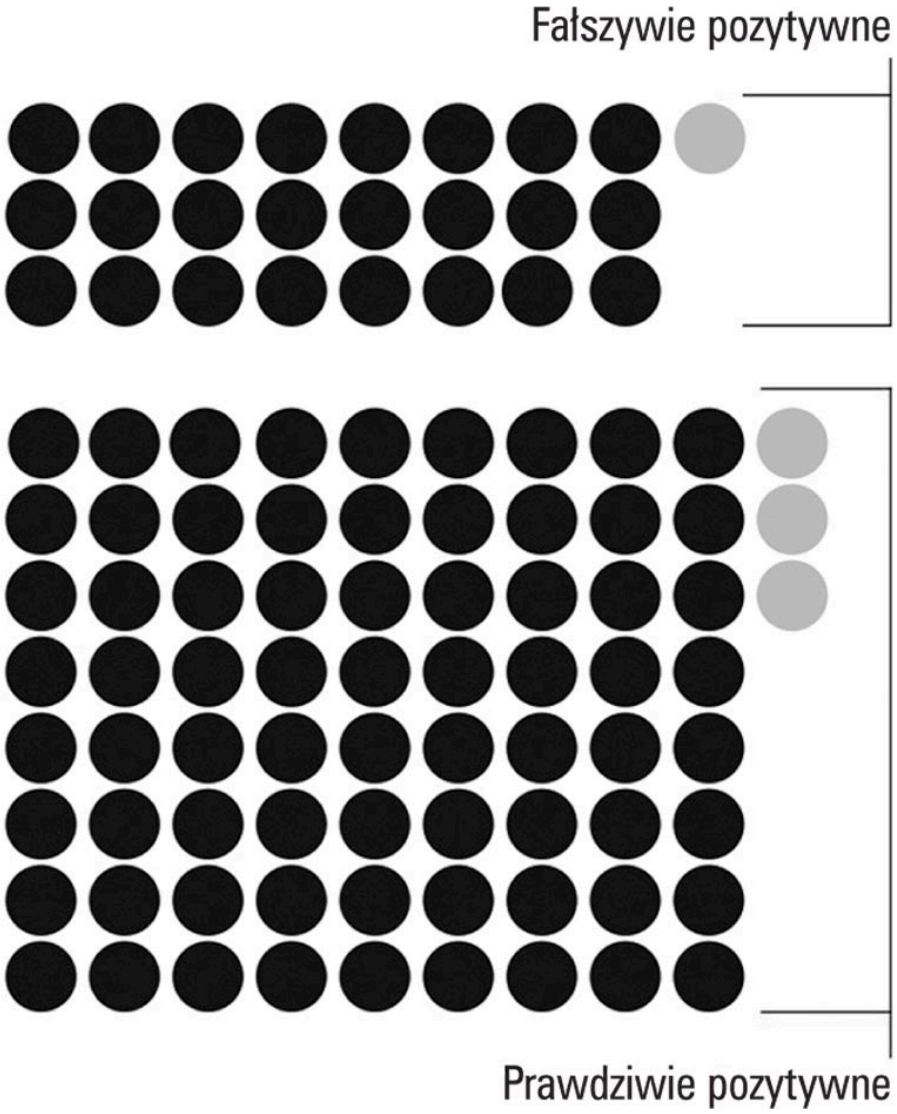
Spróbuję to wyjaśnić. Wyobraź sobie, że zaczepiasz ludzi na ulicy i korzystając z algorytmu, starasz się przewidzieć, czy dana osoba dopuści się w przyszłości zabójstwa. A ponieważ sprawcami zdecydowanej większości zbrodni są mężczyźni (ściśle mówiąc: 96 procent w skali globalnej[37]), to jeśli nasz wykrywacz morderców ma być skuteczny, siłą rzeczy musi zaliczać do grupy wysokiego ryzyka więcej mężczyzn niż kobiet.

Przyjmijmy, że nasz algorytm wykrywa zbrodniarzy z dokładnością 75 procent. Innymi słowy, trzy czwarte osób z jego grupy wysokiego ryzyka to prawdziwi Darth Vaderzy.

Po pewnym czasie uzbiera ci się grupa stu osób, które algorytm oznaczył jako potencjalnych morderców. Ponieważ algorytm opierał się na statystykach kryminalnych, 96 osób z twojej setki to mężczyźni. Pozostałe cztery to kobiety. Sytuację tę ilustruje poniższy diagram. Czarne kółka to mężczyźni, a szare to kobiety.

Ponieważ sprawdzalność prognoz algorytmu wynosi 75 procent bez względu na płeć badanych, czwarta część wytypowanych kobiet i czwarta część wytypowanych mężczyzn to nie Vaderzy, ale Skywalkerzy: to osoby niesłusznie zaliczone do grupy wysokiego ryzyka, bo w rzeczywistości nie stanowią dla nikogo zagrożenia.





A skoro już bawimy się liczbami, zauważ: z drugiego diagramu wynika, że algorytm za potencjalnych morderców uzna więcej niewinnych mężczyzn niż niewinnych kobiet – wyłącznie z tej racji, że mężczyźni popełniają więcej zbrodni niż kobiety.

Nie ma to nic wspólnego z samą zbrodnią ani z algorytmem: to po prostu matematyczny pewnik. Algorytm jest uprzedzony, bo uprzedzona jest sama rzeczywistość; więcej mężczyzn popełnia najcięższe zbrodnie i dlatego też więcej mężczyzn zostanie fałszywie posądzonych o mordercze skłonności[9*].

Dopóki odsetek osób, które dopuszczają się zbrodni, nie będzie taki sam w każdej grupie oskarżonych, dopóty stworzenie testu, który równie

dokładnie przewidywałby ryzyko dla wszystkich oraz generowałby taki sam odsetek wyników fałszywie negatywnych i fałszywie pozytywnych w każdej grupie podsądnych, jest z matematycznego punktu widzenia po prostu niemożliwe.

Afroamerykanie od stuleci zmagają się z irracjonalnymi uprzedzeniami wobec siebie i z nierównością rasową. Z tego powodu do tej pory są nadreprezentowani na niższych szczeblach drabiny społeczno-ekonomicznej, a co za tym idzie – także w policyjnych statystykach. Są również dowody na to, że za określone rodzaje przestępstw w Stanach Zjednoczonych czarnoskórzy obywatele są znacznie częściej zatrzymywani przez policję niż obywatele biali. Na przykład odsetek osób zażywających marihuanę jest taki sam w białej i czarnej populacji, a mimo to wskaźnik aresztowań wśród Afroamerykanów może być nawet ośmiokrotnie wyższy niż wśród białych[38]. Bez względu na źródła tej dysproporcji jej smutną konsekwencją jest różnica w odsetkach zatrzymań przedstawicieli tych grup etnicznych w Stanach Zjednoczonych. Czarni są też częściej aresztowani po raz drugi niż biali. Algorytm ocenia ich nie po kolorze skóry, lecz na podstawie aż nazbyt oczywistych konsekwencji głębokich, historycznie ugruntowanych podziałów występujących w społeczeństwie amerykańskim. Dopóki przedstawiciele wszystkich grup nie będą zatrzymywani równie często, dopóty tego rodzaju uprzedzenia będą po prostu matematycznie pewne.

Nie oznacza to bynajmniej odparcia wszystkich zarzutów dziennikarzy ProPubliki. Ich analiza pokazuje, z jaką łatwością algorytmy mogą odtwarzać historyczne nierówności. Nie mamy więc żadnego usprawiedliwienia dla algorytmu COMPAS. Na każdej firmie, która czerpie zyski z analizowania informacji osobistych, ciąży moralna (jeśli tymczasem jeszcze nie prawna) odpowiedzialność przyznawania się do błędów i wyjawiania zagrożeń. Tymczasem Equivant (wcześniej Northpointe), producent COMPAS-u, nadal utrzymuje szczegóły działania swojego algorytmu w najgłębszej tajemnicy, powołując się na ochronę własności intelektualnej[39].

Istnieją jednak pewne możliwości. Algorytmy nie mają cech wrodzonych, a to oznacza, że nie muszą powielać historycznych uprzedzeń. Wszystko zależy od danych, które się im ładuje. Możemy

postawić na „prostacki empiryzm” (jak to nazywa Richard Berk) i po prostu kopiować informacje do algorytmów, ale możemy też uznać, że stan rzeczy jest niesprawiedliwy i odpowiednio te informacje podkręcić.

Wrzuć do wyszukiwarki Google’a wyrażenie „profesor matematyki”. Nie zdziwi cię zapewne, że ogromna większość zdjęć dla tego hasła będzie pokazywać panów w średnim wieku stojących przy zapisanej tablicy. Kiedy ja przeprowadziłam ten test, znalazłam w pierwszej dwudziestce obrazów tylko jedną kobietę, co – przyznaję ze smutkiem – stanowi dokładne odzwierciedlenie naszej rzeczywistości: jakieś 94 procent profesorów matematyki to mężczyźni[40]. Jednak bez względu na to, jak trafne będą wyniki wyszukiwań internetowych, można by się spierać, czy tworzenie z algorytmów luster wiernie ukazujących nasz dzisiejszy świat jest takie dobre – zwłaszcza gdy lustro odzwierciedla stan rzeczy będący wyłącznie rezultatem stuleci uprzedzeń. Google mógłby więc pójść inną drogą i subtelnie podkręcić swój algorytm, żeby pokazując zdjęcia profesorów matematyki, dawać pierwszeństwo zdjęciom kobiet i osób o kolorze skóry innym niż biały – żeby bardziej zrównoważyć szale i pokazywać nam raczej takie społeczeństwo, do jakiego dążymy, niż to, w którym obecnie żyjemy.

Podobnie jest z wymiarem sprawiedliwości. Zanim skorzystamy z jakiegoś algorytmu, zadajmy sobie pytanie: jaki procent osób z danej grupy spodziewalibyśmy się uznać za osoby wysokiego ryzyka w idealnie sprawiedliwym i równym społeczeństwie? Algorytm daje nam możliwość przejścia na skróty prosto do tej liczby. Gdybyśmy jednak doszli do wniosku, że usunięcie z wymiaru sprawiedliwości wszelkich uprzedzeń za jednym zamachem nie jest najlepszym rozwiązaniem, moglibyśmy kazać algorytmowi podążać stopniowo w tym kierunku.

Mamy również pole manewru traktowania podsądnych o wysokim ryzyku recydywy. Na przesłuchaniach w sprawie zwolnienia za kaucją, gdzie stopień prawdopodobieństwa niestawienia się oskarżonego w wyznaczonym terminie przed sądem jest kluczowym elementem prognozy algorytmu, normą jest odmawianie zwolnienia z aresztu wszystkim osobom zakwalifikowanym do grupy wysokiego ryzyka. A przecież algorytm mógłby brać pod uwagę również powody, dla których oskarżony nie zjawił się na swoją rozprawę. Czy ma dostęp do odpowiednich środków komunikacji, żeby dojechać do sądu? Może nie ma

nikogo, kto mógłby zająć się dziećmi w dniu rozprawy? Czy nie można by tak zaprogramować algorytmu, żeby potencjalne nierówności społeczne starał się niwelować, zamiast je jeszcze bardziej pogłębiać?

Odpowiedzi na te pytania powinny paść jednak raczej na forum otwartej debaty publicznej i w gabinetach rządowych, a nie na posiedzeniach zarządu prywatnych przedsiębiorstw. Na szczęście coraz wyraźniej słychać głosy wzywające do powołania organu regulującego użycie algorytmów i kontrolującego branżę. Na takiej samej zasadzie, na jakiej amerykańska Agencja Żywności i Leków reguluje rynek farmaceutyków, organ taki sprawdzałby za zamkniętymi drzwiami dokładność, konsekwencję i uprzedzenia algorytmów i miałby prerogatywy pozwalające na dopuszczenie bądź niedopuszczenie do stosowania danego produktu w salach sądowych. Dopóki takiego organu nie ma, nie do przecenienia są organizacje pokroju ProPublica, które zaczynają patrzeć algorytmom na ręce – choć wystrzegałabym się przed nawoływaniem do całkowitego zakazu ich stosowania, zanim spokojnie zastanowimy się nad tym, jak nasz wymiar sprawiedliwości będzie jutro wyglądać bez nich.

Trudne decyzje

To niezwykle ważna kwestia, którą musimy poruszyć. Jeżeli zakażemy algorytmów, co pozostanie z naszego sądownictwa? Pytanie jest zasadne, ponieważ wykazano, że brak jednolitości orzeczeń to wcale nie jedyna przypadłość, na którą cierpią sędziowie.

Prawo mówi, że rasa, płeć i klasa nie mogą wpływać na decyzje sędziów. (W końcu sprawiedliwość powinna być ślepa). A jednak – mimo że ogromna większość sędziów stara się zachować bezstronność, wielokrotnie udowodniono, że w istocie podsądni są przez nich dyskryminowani. Z amerykańskich statystyk wynika, że czarnoskórzy oskarżeni idą do więzienia na dłużej[41], mają mniejsze szanse na zwolnienie z aresztu za kaucją[42] i większe na dostanie kary śmierci[43]; jest też bardziej prawdopodobne, że kiedy już trafią do celi śmierci, wyrok na nich zostanie wykonany[44]. Inne badania wykazały, że mężczyźni są karani bardziej surowo niż kobiety za tę samą zbrodnię[45] i że podsądni

o niskim wykształceniu i poziomie dochodów dostają znacznie dłuższe kary pozbawienia wolności[46].

Sytuacja jest analogiczna do już omawianego przypadku uprzedzeń algorytmów – tak nierówne traktowanie podsądnych jest nie tyle rezultatem umyślnego działania, ile bezmyślnego powielania historii. Po prostu uprzedzenia społeczne i kulturowe mogą automatycznie dochodzić do głosu w procesie podejmowania decyzji przez człowieka.

Aby wyjaśnić, dlaczego tak się dzieje, musimy najpierw poznać kilka prostych faktów związanych z ludzką intuicją. Opuśćmy więc na chwilę salę sądową i przyjrzyjmy się następującemu zadaniu:

> Kij do bejsbola i piłka kosztują razem 1 dolara i 10 centów.
> Kij kosztuje o dolara więcej niż piłka.
> Ile kosztuje piłka?

Zagadka ta, którą zamieścił w swoim bestsellerze *Pułapki myślenia. O myśleniu szybkim i wolnym*[47] ekonomista i psycholog, laureat Nagrody Nobla Daniel Kahneman, dobrze ilustruje pułapkę, w którą wszyscy wpadamy, kiedy zaczynamy myśleć.

Zagadka ma proste rozwiązanie, na które łatwo wpaść po chwili namysłu – ale ma także odpowiedź błędną – właśnie tę, która wprost nam się narzuca. Ta błędna odpowiedź, o której od razu myślimy, to oczywiście – 10 centów[10*].

Nie rób sobie wyrzutów, jeżeli nie udzieliłeś właściwej odpowiedzi (5 centów). Ten sam błąd popełniło 71,8 procent sędziów, którym dano tę zagadkę[48]. Nawet ci z nich, którzy podali ostatecznie właściwe rozwiązanie, musieli oprzeć się pokusie, żeby posłuchać swojej pierwszej myśli.

To właśnie rywalizacja intuicji z namysłem stanowi sedno naszej historii o procesie podejmowania decyzji przez sędziów. Psychologowie zgadzają się na ogół, że istnieją dwa rodzaje myślenia, dwa systemy. System 1 jest automatyczny, instynktowny, lecz podatny na błędy. (To on podpowiadał ci

10 centów w przytoczonej wyżej zagadce). System 2 jest powolny, analityczny, rozważny, często jednak bardzo leniwy[49].

Kiedy spytasz kogoś, co skłoniło go do podjęcia określonej decyzji, odpowiedź wyartykułuje System 2, który – słowami Kahnemana – „formułuje oceny i dokonuje wyborów, jednak często akceptuje albo racjonalizuje myśli i uczucia wygenerowane przez System 1”[50].

Pod tym względem sędziowie niczym nie różnią się od nas. Ostatecznie też są ludźmi i padają czasem ofiarą kaprysów czy słabości, jak my wszyscy. Tak to już jest, że nasze mózgi po prostu nie są przystosowane do długotrwałego, racjonalnego analizowania złożonych problemów. Niełatwo nam rozważać wszystkie za i przeciw, żeby potem połączyć wszystko w spójną całość; cały czas musimy odpierać ataki Systemu 1, który namawia nas do pójścia na poznawcze skróty.

Życzylibyśmy sobie na przykład, żeby sędziowie na przesłuchaniach w sprawie zwolnienia za kaucją potrafili ogarnąć myślą całą sprawę, żeby starannie rozważyli wszystkie za i przeciw przed podjęciem decyzji. Niestety, psychologowie wykazali, że jest inaczej – cała strategia sędziowska to nic innego, jak przeglądanie i odhaczanie w myślach listy kluczowych elementów. Jeśli podczas rozpatrywania sprawy oskarżonego przy którymś z tych elementów zapali się lampka ostrzegawcza – wcześniejsze wyroki, naciski społeczne, prośba ze strony oskarżenia – sędzia zamyka sprawę i wydaje decyzję odmowną[51].

Problem w tym, że bardzo wiele z tych lampek ostrzegawczych podłączonych jest do kategorii rasy, płci i poziomu wykształcenia. Sędziowie nie są świadomi tego, że słuchają podszeptów intuicji uważniej, niż powinni – czym mimowolnie przyczyniają się do powielania zakodowanych w systemie uprzedzeń.

A to jeszcze nie wszystko – to zaledwie czubek góry lodowej. Niestety – ludzie są beznadziejnie nieudolni, gdy trzeba wydać sprawiedliwy i uczciwy sąd.

Jeżeli zdołaliście kiedyś przekonać samych siebie, że bardzo droga rzecz ma sensowną cenę tylko dlatego, że została przeceniona o połowę (ja robię to regularnie), to wiesz już wszystko o tak zwanym efekcie zakotwiczenia. Nie radzimy sobie z przykładaniem miar liczbowych do przedmiotów;

znacznie wygodniej jest nam porównywać wartości, niż wyciągać z rękawa konkretne wielkości. W reklamie z efektu zakotwiczenia korzysta się od lat – nie tylko po to, żeby kazać nam cenić sobie wysoko pewne rzeczy, ale także po to, by wpływać na ilość kupowanych przez nas produktów. Weźmy tablice w supermarkecie, na których czytamy na przykład: „Limit 12 puszek zupy na jednego klienta". Nie mają one bynajmniej zapobiec wykupieniu całej partii towaru przez amatorów puszkowanej zupy, lecz subtelnie modyfikować nasze wyobrażenia o tym, ile takich puszek moglibyśmy potrzebować. Mózg kotwiczy się na liczbie 12 i szuka czegoś poniżej. Badanie z lat dziewięćdziesiątych ubiegłego wieku pokazało, że tego typu informacja może zwiększyć średnią sprzedaż puszek na jednego konsumenta z 3,3 do 7[52].

Spodziewam się, iż nie zaskoczy cię informacja, że to działa także na sędziów: są bardziej skłonni przyznawać większe odszkodowania, kiedy oskarżyciel domaga się wyższych kwot[53], i orzekać dłuższe wyroki pozbawienia wolności, jeśli oskarżyciel żąda surowszej kary[54]. W jednym badaniu wykazano nawet, że można znacząco wpłynąć na długość orzekanej kary więzienia w hipotetycznej sprawie, każąc dziennikarzowi zatelefonować do sędziego podczas przerwy i subtelnie wpleść sugerowaną karę do rozmowy. („Czy pańskim zdaniem wyrok w tym procesie powinien być większy czy mniejszy od trzech lat?")[55]. A chyba najgorsze z tego wszystkiego jest to, że prawdopodobnie można wpłynąć na decyzję sędziego, nakłoniwszy go tuż przed rozpatrzeniem sprawy do rzucenia kostką do gry[56]. Ten sposób manipulacji działał nawet na najbardziej doświadczonych sędziów[57].

W procesie porównywania liczb ludzie popełniają jeszcze jeden błąd, który wpływa na uczciwość i bezstronność orzeczeń sądowych. Być może sami spostrzegliście u siebie tę dziwną przypadłość umysłu. Skutek podkręcenia głośności w twojej wieży o jeden ząbek jest tym słabszy, im głośniej gra muzyka; skok cenowy z 1,20 do 2,20 funtów wydaje się ogromny, podczas gdy wzrost ceny z 67 na 68 funtów nie ma jakby większego znaczenia; czas też biegnie coraz szybciej w miarę, jak się starzejemy. Dzieje się tak dlatego, że ludzkie zmysły funkcjonują raczej w oparciu o wielkości względne niż wartości absolutne. Rok nie stanowi

dla nas określonej miary czasu; każdy kolejny przeżywamy jako coraz mniejszą porcję przeżytego do tej pory życia. Wymiary porcji czasu, pieniędzy czy głośności stosują się do bardzo prostej formuły matematycznej znanej jako prawo Webera.

W najprostszym ujęciu prawo Webera mówi, że najmniejsza zmiana bodźca, jaką jesteśmy w stanie zarejestrować – tak zwana najmniejsza zauważalna różnica – jest proporcjonalna do bodźca początkowego. Nic dziwnego, że także to odkrycie zostało wykorzystane w marketingu. Specjaliści od sprzedaży doskonale wiedzą, jak długo będzie im uchodzić na sucho zmniejszanie rozmiaru tabliczki czekolady, zanim klienci spostrzegą zmianę, albo o ile dokładnie da się podkręcić cenę produktu, zanim kupujący dojdzie do wniosku, że lepiej jednak zajrzeć do konkurencji.

W kontekście wymiaru sprawiedliwości prawo Webera wpływa na długość odsiadki, jaką orzekają sędziowie. Im surowsza kara grozi oskarżonemu, tym większe stają się przedziały między możliwymi wyrokami; jeżeli przestępstwo jest marginalnie cięższe od tego zagrożonego karą 20 lat pozbawienia wolności, dodatkowe 3 miesiące to jakby za mało: człowiek ma wrażenie, że między 20 latami a 20 latami i 3 miesiącami nie ma praktycznie żadnej różnicy. A przecież to nieprawda; różnica jest, i to znaczna: trzy miesiące więzienia nie staną się ani trochę krótsze tylko dlatego, że wcześniej było ich dużo więcej. A jednak zamiast dodać do wyroku kilka miesięcy, sędziowie przeskakują od razu parę lat, do wyraźnie dłuższej kary, czyli na przykład do 25 lat pozbawienia wolności[58].

Wiemy, że tak to działa, bo możemy porównać długość faktycznie orzeczonych kar z tymi przewidywanymi przez prawo Webera. Autorzy badania, którzy w 2017 roku zweryfikowali ponad sto tysięcy orzeczeń z Wielkiej Brytanii i Australii, stwierdzili, że 99 procent skazanych dostało wyrok zgodny z tą formułą[59].

„I to bez względu na to, jakiego przestępstwa się dopuściłaś, jakiego rodzaju jesteś oskarżoną, w którym kraju zostałaś skazana i czy w grę wchodziła kara pozbawienia, czy też ograniczenia wolności”, powiedziała mi Mandeep Dhami, główna autorka pracy. O wszystkim decydowała

liczba, która przyszła sędziemu do głowy i wydała się mniej więcej odpowiednia.

Niestety, o stronniczych sędziach mogłabym pisać jeszcze bardzo długo. Sędziowie mający córki są bardziej skłonni do wymierzania kobietom łagodniejszych kar[60] – i mniej skłonni do zwalniania podsądnych za kaucją wkrótce po przegranej swojej drużyny sportowej. Autorzy pewnego słynnego już badania zasugerowali nawet, że na prawdopodobieństwo orzeczenia łagodniejszej kary wpływa pora dnia[61]. Z kolei inny eksperyment[62] (choć musi on zostać jeszcze powtórzony, nadal też trwa debata o stopniu jego ewentualnego wpływu) pokazuje, że najgorzej jest mieć rozprawę tuż przed przerwą na lunch: okazało się, że sędziowie byli najbardziej skłonni do zwolnienia oskarżonego za kaucją zaraz po powrocie z przerwy i najmniej do tego skłonni, kiedy zbliżała się pora obiadowa.

Jeszcze inne badanie pokazało, że sędzia będzie unikał wydawania zbyt wielu takich samych orzeczeń pod rząd. Dlatego też twoje szanse wyjścia z aresztu za poręczeniem majątkowym spadają dramatycznie, jeśli przed twoją rozprawą sędzia miał już cztery podobne sprawy i wszystkich podsądnych wypuścił za kaucją[63].

Część badaczy twierdzi również, że to, jak postrzegamy nieznajomych, zależy od temperatury trzymanej w rękach szklanki z napojem. Podobno jeżeli tuż przed spotkaniem z obcą osobą ktoś wręczy nam coś ciepłego do picia, jest bardziej prawdopodobne, że nieznajomy wyda nam się człowiekiem cieplejszym, bardziej wspaniałomyślnym i troskliwym[64].

Ta długa lista obejmuje jedynie te czynniki, które możemy zmierzyć. Bez wątpienia istnieje też ogromna liczba innych zmiennych, które subtelnie wpływają na nasze zachowanie, lecz w warunkach sali sądowej nie da się ich ująć w ramy eksperymentu.

Podsumowanie

Będę z tobą szczera. Kiedy pierwszy raz usłyszałam, że w salach sądowych używa się algorytmów, byłam temu przeciwna. Uważałam, że skoro algorytmy się mylą, a ich pomyłka może oznaczać, że człowiek zostanie

pozbawiony wolności, oddawanie tak wielkiej władzy maszynie jest nieodpowiedzialne.

Nie byłam w tym rozumowaniu odosobniona. Moje odczucia podziela wiele osób (być może nawet większość), które weszły w kolizję z prawem karnym. Mandeep Dhami powiedziała mi, co przestępcy, z którymi pracowała, sądzą o sposobie rozstrzygania o ich przyszłości: „Chociaż mieli świadomość, że sędzia może popełnić więcej błędów, woleli jego od algorytmu. Chcieli mieć gwarancję, że nad tym wszystkim czuwa jednak człowiek".

Tak samo zresztą myślą prawnicy. Jeden z londyńskich adwokatów wyjawił mi w rozmowie, że jego zadanie polega właśnie na wykorzystywaniu tej istniejącej w systemie niepewności, co z algorytmami byłoby już trudniejsze. „Im bardziej przewidywalne stają się decyzje, tym mniej miejsca pozostaje na sztukę adwokacką".

Kiedy jednak spytałam Mandeep Dhami, co ona by myślała, gdyby sama miała trafić do więzienia, przedstawiła mi w tej sprawie zupełnie inną opinię:

„Nie chciałabym, żeby osoba, która będzie decydować o mojej przyszłości, polegała wyłącznie na swojej intuicji. Chciałabym, żeby ktoś bazował na racjonalnej strategii. Bronimy tej naszej dyskrecjonalności sędziów, jakby to była jakaś święta krowa. Jakby to było nie wiadomo co. Mimo że badania pokazują, że wcale tak nie jest. Jest dokładnie odwrotnie".

Tak jak wszyscy, ja również uważam, że sędziowie powinni być na tyle bezstronni, na ile to możliwe. W swoich decyzjach powinni się kierować czynami podsądnego, a nie jego przynależnością do tej czy innej społeczności lub grupy. Pod tym względem algorytmy nie wypadają zbyt dobrze. Nie wystarczy jednak po prostu wyliczyć wady komputerów; nie chodzi tu przecież o wybór między niedoskonałą maszyną a niedoścignionym, idealnym wzorcem. Jedynym uczciwym porównaniem będzie zestawienie dwóch systemów: takiego, który korzysta z algorytmów, i takiego, którym byśmy dysponowali, gdybyśmy zrezygnowali z ich stosowania.

Im więcej czytałam, im więcej odbyłam spotkań ze specjalistami, tym bardziej byłam przekonana, że trochę za wysoko stawiamy sędziom

poprzeczkę. Niesprawiedliwość jest wbudowana w nasze ludzkie układy. Na każdego Christophera Drew Brooksa niesprawiedliwie potraktowanego przez algorytm przypadają dziesiątki Nicholasów Robinsonów pogrążonych przez sędziów bez pomocy komputera. Wprowadzenie nawet niedoskonałego algorytmu, który będzie korygował sędziów w ich często fałszywym oglądzie rzeczywistości, jest moim zdaniem krokiem we właściwym kierunku. Dobrze zaprojektowany i właściwie wyregulowany algorytm pomoże nam wyrugować wbudowane w system uprzedzenia oraz przypadkowe błędy. Nie zdołamy zmienić sposobu postępowania wszystkich sędziów od razu – tym bardziej że jej przedstawiciele nie są w stanie powiedzieć nam nawet, w jaki sposób podejmują decyzje.

Jeśli chcemy zaprojektować algorytm wspierający sędziów, musimy dobrze się zastanowić, czego dokładnie oczekujemy po naszym wymiarze sprawiedliwości. Nie możemy przymykać oczu na niedoskonałości i mieć nadzieję, że wszystko się jakoś ułoży; algorytmy wymagają klarownej i jednoznacznej odpowiedzi na pytanie, co dokładnie mają robić, oraz solidnej wiedzy o ludzkich słabościach, które mają wyeliminować. To zaś zmusza nas do podjęcia trudnej debaty o tym, jak dokładnie powinny być podejmowane decyzje na sali sądowej. Nie będzie to łatwe, ale tylko tak zdołamy ustalić, czy maszyna kiedykolwiek spełni nasze wymagania.

Tarcia i napięcia w systemie sądowniczym tylko mącą wodę i niezwykle utrudniają udzielenie odpowiedzi na powyższe pytania. Są jednak inne pola, powoli penetrowane przez algorytmy, gdzie decyzje nie powodują aż tak zaciekłych sporów, a zadania algorytmów i ich pozytywny wkład w życie społeczne są bardziej oczywiste.

Medycyna

W 2015 roku grupa naukowców przeprowadziła niezwykłe, pionierskie studium diagnozowania raka[1]. Badacze dali monitor dotykowy szesnastu testerom i poprosili ich o posortowanie skanów tkanki piersiowej pobranej w wyniku biopsji od prawdziwych pacjentek. Próbki patologiczne pocięto na cienkie skrawki. Chemicznie zabarwiono naczynia krwionośne i przewody mleczne na czerwono, purpurowo i niebiesko. Jedynym zadaniem testerów było ustalenie, czy barwne wzory na skanach wskazują na ukrytego między komórkami raka.

Po krótkim treningu testerzy przystąpili do dzieła – i osiągnęli imponujący wynik. Pracując niezależnie, poprawnie ocenili 85 procent próbek.

Naukowcy spostrzegli jednak coś zdumiewającego. Kiedy zaczęli sumować odpowiedzi pojedynczych testerów, żeby sprawdzić ogólną ocenę każdego skanu – dokładność osiągnęła astronomiczny wynik 99 procent.

To jednak nie umiejętności testerów stanowiły o wyjątkowości owego eksperymentu, lecz to, kim byli nasi dzielni diagności. Nie byli bowiem onkologami. Nie byli patologami. Nie byli pielęgniarkami. Nie byli nawet studentami medycyny. Testerami były gołębie.

Patologowie nie muszą się jeszcze bać o swoją przyszłość – nie sądzę, żeby nawet naukowcy, którzy zaprojektowali ten eksperyment, chcieli zasugerować, że lekarzy powinno się zastąpić pospolitym ptactwem. Doświadczenie udowodniło jednak bardzo ważną hipotezę: dostrzeganie wzorców ukrywających się pośród skupisk komórek nie jest umiejętnością wyłącznie ludzką. A zatem – jeśli poradził z tym sobie gołąb, to czemu nie miałby algorytm?

Łowcy wzorców

Cała historia i praktyka współczesnej medycyny opiera się na wyszukiwaniu prawidłowości w zbiorach danych. Odkąd mniej więcej dwa i pół tysiąca lat temu w starożytnej Grecji Hipokrates założył swoją szkołę medycyny, obserwacja, eksperyment i analiza informacji stały się podstawowymi elementami walki o nasze zdrowie.

Wcześniej medycyna była w zasadzie tożsama z magią. Ludzie wierzyli, że choroba jest karą za rozgniewanie jakiegoś bóstwa lub skutkiem opanowania ciała przez złego ducha. Na praktyki ówczesnych „lekarzy" składały się zatem głównie zawodzenia i zaklęcia, co dzisiaj brzmi może zabawnie, ale umierającemu człowiekowi, który pokładał w tych metodach całą swoją nadzieję, zapewne nie było do śmiechu.

Co prawda Hipokrates nie od razu wyleczył świat z irracjonalności i zabobonu (według jednej z pogłosek miał za córkę trzydziestometrowego smoka)[2], niemniej jego podejście do medycyny było naprawdę rewolucyjne. Wielki Grek uważał, że przyczyn choroby można dociec w drodze racjonalnej obserwacji, a nie za pomocą magii. Uczyniwszy z opisu przypadku i obserwacji fundament swojej sztuki, Hipokrates stworzył naukę medyczną – nic więc dziwnego, że cieszy się dziś reputacją „ojca współczesnej medycyny"[3].

Mimo że „naukowość" wywodów Hipokratesa i jego kolegów niezupełnie odpowiada współczesnym standardom (starożytni wierzyli, że zdrowie to równowaga krwi, flegmy, żółci i czarnej żółci)[4], to jednak wnioski, jakie wyciągali z zebranych informacji, z pewnością już tak[5]. (To oni jako pierwsi przekazali nam, między innymi, takie spostrzeżenie: „Pacjenci z natury swej otyli zazwyczaj umierają wcześniej od tych szczupłych"). Jest to temat, który powraca od tysiącleci. Choć nasze rozumienie nauki kilkakrotnie poszło w niewłaściwą stronę, dzięki zdolności do odnajdywania wzorców, klasyfikowania symptomów i wykorzystywania tych obserwacji do określania rokowań robimy ciągłe postępy.

Historia medycyny zna mnóstwo przykładów. Weźmy piętnastowieczne Chiny, gdzie lekarze po raz pierwszy uzmysłowili sobie, że można uodpornić ludzi na czarną ospę. Po kilku wiekach eksperymentów

Chińczycy odkryli prawidłowość, która pozwoliła zredukować ryzyko śmierci z powodu choroby aż dziesięciokrotnie: wystarczyło znaleźć zakażonego z lekkim przebiegiem choroby, pobrać od niego strupy, wysuszyć je i sproszkować, a następnie wdmuchnąć do nosa osoby zdrowej[6].

Albo złota era sztuki lekarskiej w XIX wieku, kiedy medycyna przejmowała coraz więcej z metodologii nauk i szukanie prawidłowości w danych stało się podstawowym narzędziem pracy lekarzy. Takich jak Węgier Ignaz Semmelweis, który w latach czterdziestych XIX wieku zauważył coś uderzającego w okolicznościach zgonów okołoporodowych: kobiety, które rodziły w izbach prowadzonych przez lekarzy, były pięciokrotnie bardziej narażone na sepsę niż kobiety w izbach prowadzonych przez akuszerki. Dane podsuwały natychmiastowe rozwiązanie tej zagadki: lekarze przystępowali do rodzących kobiet od razu po wyjściu z prosektorium, bez mycia rąk[7].

Ta sama zasada, której trzymali się lekarze w piętnastowiecznych Chinach i dziewiętnastowiecznej Europie, obowiązuje dziś na całym świecie. Nie tylko podczas badania chorób całej populacji, lecz także w codziennej praktyce lekarza pierwszego kontaktu. Czy kość jest złamana czy nie? Czy ten ból głowy jest całkowicie normalny, czy też może być oznaką czegoś znacznie groźniejszego? Czy zapisanie antybiotyku spowoduje zniknięcie tego czyraka? Wszystkie te pytania wiążą się z rozpoznawaniem prawidłowości, klasyfikowaniem i przewidywaniem. A tak się składa, że w tym algorytmy są naprawdę dobre.

Oczywiście jest wiele aspektów pracy lekarza, w których algorytmy prawdopodobnie nigdy nie zdołają zastąpić człowieka. Przede wszystkim – empatia. Albo zdolność wspierania pacjentów zmagających się z trudnościami społecznymi, psychologicznymi czy nawet finansowymi. Niemniej są takie obszary medycyny, w których algorytmy mogą nas zastąpić. Zwłaszcza w tych rolach, w których spotykamy się z rozpoznawaniem prawidłowości w najczystszej postaci, gdzie klasyfikowanie i prognozowanie stają się tak ważne, że cała reszta przestaje się liczyć niemal zupełnie. A więc szczególnie w takich dziedzinach jak patologia.

Patologowie to lekarze, których pacjenci spotykają rzadko. Kiedy próbka twojej krwi czy tkanki zostaje wysłana do zewnętrznego laboratorium, to właśnie patolog będzie ją badał i opisywał. Patolog to ostatnie ogniwo długiej linii diagnostycznej, gdzie wiedza, precyzja i wiarygodność są nie do przecenienia. To często on orzeka, czy masz raka, czy nie. A kiedy wysłana do analizy próbka stanowi jedyne ogniwo pośrednie między tobą a chemioterapią, operacją czy czymś jeszcze gorszym, to chyba oczywiste, że wszyscy chcemy mieć pewność, że patolog się nie pomylił.

A jego praca nie jest wcale łatwa. Przeciętny patolog musi dziennie sklasyfikować setki skrawków zawierających dziesiątki, a niekiedy setki tysięcy komórek ściśniętych między maleńkimi szkiełkami – jest to dosłownie szukanie igły w stogu siana. Każdą próbkę trzeba systematycznie i niezwykle dokładnie obejrzeć, bo gdzieś w tej wielkiej galaktyce komórek widocznych w okularze mikroskopu może się ukrywać maciupeńka anomalia.

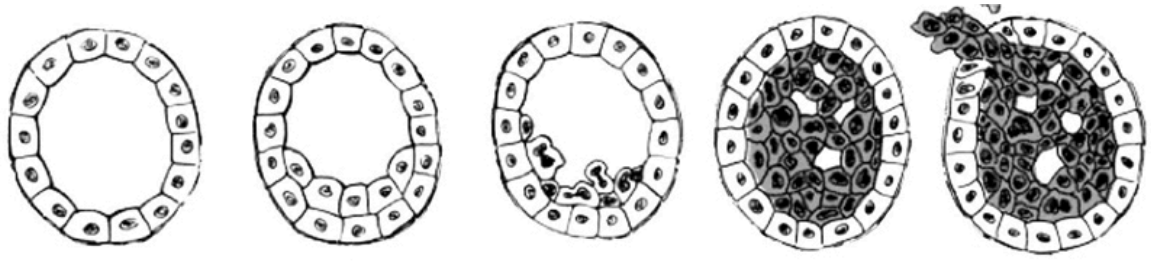
„To zadanie tak trudne, że praktycznie niemożliwe", mówi Andy Beck[8], patolog z Harvardu, założyciel firmy o nazwie PathAI, która od 2016 roku tworzy algorytmy klasyfikujące materiał biopsyjny. „Gdyby patolog mógł dokładnie obejrzeć pięć skrawków dziennie, przypuszczalnie osiągnąłby doskonałość. Ale w rzeczywistym świecie tak nie jest".

Nie jest tak na pewno. W rzeczywistym świecie pracę patologa komplikują jeszcze frustrujące zawiłości biologii. Wróćmy do przykładu z rakiem piersi, którego tak dobrze rozpoznawały gołębie. Orzeczenie, czy pacjentka jest chora, nie jest kwestią udzielenia prostej, zerojedynkowej odpowiedzi: tak lub nie. Diagnozy raka piersi tworzą pewne spektrum. Na jednym biegunie są próbki łagodne, na których normalne komórki wyglądają dokładnie tak, jak wyglądać powinny. Na drugim są najpaskudniejsze rodzaje guzów – inwazyjne formy choroby, w których komórki rakowe porzuciły przewody mleczne i zaczęły naciekać otaczającą je tkankę. Przypadki z obu tych biegunów spektrum rozpoznać jest stosunkowo łatwo. Niedawno przeprowadzony eksperyment wykazał, że patologowie poprawnie diagnozują 96 procent prostych postaci raka

złośliwego, czyli osiągają z grubsza taką samą trafność jak stado gołębi, którym dano podobne zadanie[9].

Pomiędzy tymi skrajnościami – przypadkami całkowicie normalnymi i już na pierwszy rzut oka nadzwyczaj złośliwymi – kryje się jednak kilka innych, bardziej dyskusyjnych kategorii, co pokazują poniższe rysunki. Twoja próbka mogłaby zawierać zbitkę nietypowych komórek, które wyglądają trochę podejrzanie, lecz niekoniecznie oznaczają kłopoty. Mogłabyś też mieć zmiany przednowotworowe, które mogą, choć wcale nie muszą, okazać się poważne. No i mogłabyś mieć raka, który jeszcze nie wyszedł poza przewody mleczne (jest to tak zwany rak przewodowy *in situ*).

To, do której z tych kategorii zostanie przypisana twoja próbka, prawdopodobnie znacznie zaważy na przebiegu leczenia. W zależności od lokalizacji przypadku w spektrum lekarz może zalecić cokolwiek – od mastektomii do całkowitej rezygnacji z interwencji medycznej.



Normalny przewód | Rozrost przewodowy | Atypowy rozrost przewodowy | Rak przewodowy *in situ* | Rak przewodowy inwazyjny

Kłopot w tym, że odróżnienie od siebie niejednoznacznych kategorii spektrum może być niewiarygodnie trudne. Nawet eksperci w dziedzinie patologii mogą się nie zgadzać co do poprawnej diagnozy pojedynczej próbki. Żeby zbadać, jak bardzo różnią się od siebie opinie lekarzy, autorzy pewnego studium z 2015 roku przedstawili 115 patologom 72 próbki biopsyjne tkanki piersiowej uznane za wykazujące łagodne anomalie (czyli należące do kategorii mniej więcej ze środka naszego spektrum) i poprosili o ich opis. Wynik był niepokojący – lekarze postawili tę samą diagnozę tylko w 48 procentach przypadków[10].

Kiedy przy stawianiu diagnozy spadamy do poziomu 50 na 50, równie dobrze moglibyśmy rzucać monetą. Orzełek – i dostaje ci się zupełnie niepotrzebna mastektomia (co w Stanach Zjednoczonych kosztuje setki tysięcy dolarów). Reszka – i być może straciłaś szansę na udaną interwencję na najwcześniejszym etapie rozwoju raka. Tak czy owak, skutki mogą być tragiczne.

Kiedy stawka jest tak wysoka, trafność diagnozy to rzecz najważniejsza. A zatem – czy algorytm dałby radę poprawić ten wynik?

Maszyny, które widzą

Jeszcze nie tak dawno stworzenie algorytmu umiejącego rozpoznać cokolwiek na zdjęciu – nie mówiąc o komórkach rakowych – uważano za zadanie beznadziejnie trudne. Pojmowanie obrazów jest dla nas czymś naturalnym i prostym, ale spróbuj wytłumaczyć, co dokładnie nam to umożliwia – staniesz przed kolejnym niewyobrażalnie trudnym problemem.

Aby zrozumieć jego skalę, wyobraź sobie, że piszesz dla komputera instrukcje rozpoznawania psa na zdjęciu. Zaczęłabyś pewnie od rzeczy oczywistych: pies ma cztery łapy, lekko zagięte uszy, sierść i tak dalej. Ale co z obrazkami, na których zwierzę siedzi? I nie widać wszystkich łap? A co z psami, które mają uszy spiczaste? Albo zupełnie oklapnięte? Co z psami, które nie są zwrócone przodem w stronę obiektywu? I czym się różni „sierść" od puszystego dywanu? Albo sierści owcy? Lub trawy?

Oczywiście dałoby się to wszystko załatwić dodatkowymi instrukcjami opisującymi każdy możliwy typ psiego ucha, sierści czy sposobu siedzenia, tylko że wtedy twój algorytm rozrósłby się szybko do takich rozmiarów, że przestałby funkcjonować, zanim w ogóle zaczęłabyś odróżniać psy od innych czworonożnych zwierząt pokrytych sierścią. Musisz znaleźć inny sposób – odejść od paradygmatu opartego na instrukcjach i wykorzystać tak zwaną sieć neuronową[11].

Możesz sobie wyobrazić sieć neuronową jako olbrzymią konstrukcję matematyczną wyposażoną w całą masę pokręteł i suwaków. Wrzucasz swoje zdjęcie do maszyny, która analizuje je i podaje wynik będący mniej lub bardziej pewną odpowiedzią na pytanie, co zdjęcie przedstawia;

maszyna wyraża ją stopniem prawdopodobieństwa dla każdej kategorii: pies, niepies.

Na początku twoja sieć neuronowa to zwykła kupa złomu: kiedy zaczyna od poziomu zerowego, nie ma jeszcze bladego pojęcia o tym, co jest, a co nie jest psem. Wszystkie pokrętła i suwaki są ustawione byle jak, i tak też algorytm odpowiada – na tym etapie nie potrafiłby poprawnie rozpoznać niczego nawet pod groźbą odłączenia od prądu. Ale z każdym kolejnym zdjęciem zaczynasz regulować pokrętła i suwaki, powoli ucząc swoją maszynę.

Kolejne zdjęcie psa. Po każdej prognozie maszyna uruchamia zbiór reguł matematycznych, które dostrajają wszystkie pokrętła i suwaki, żeby prognoza była bliższa właściwej odpowiedzi. Wrzucasz do maszyny następne zdjęcie i następne, regulując ją; kiedy popełni błąd – rozjaśniasz ścieżki prowadzące do celu i przyciemniasz te wiodące na manowce. Informacja o tym, co upodabnia jedno zdjęcie psa do drugiego, przechodzi zwrotnie przez całą sieć. I tak przez setki i tysiące zdjęć, dopóki trafność prognoz algorytmu nie osiągnie pewnego stałego poziomu. W końcu będzie można pokazać maszynie zdjęcie, którego nigdy wcześniej nie widziała, a odpowie ci z dużym prawdopodobieństwem, czy jest na nim pies, czy go nie ma.

Zaskakujące w sieciach neuronowych jest to, że ich operatorzy zazwyczaj nie rozumieją, jak i dlaczego algorytm dochodzi do swoich wniosków. Sieć neuronowa klasyfikująca zdjęcia psów nie działa na cechach, które ty czy ja moglibyśmy uznać za psokształtne. Nie szuka stopnia „szkockiej owczarkowości" ani „niemieckiej dogowatości" – jego funkcjonowanie jest znacznie bardziej abstrakcyjne: algorytm skupia się na wychwytywaniu na obrazach wzorców krawędzi oraz światła i cienia, które dla nas nie mają większego sensu (zajrzyj do przykładu z rozpoznawaniem obrazów w rozdziale *Władza*, a zrozumiesz, co mam na myśli). Nie jesteśmy w stanie tego procesu skonceptualizować i dlatego operator algorytmu może wiedzieć jedynie tyle, że wyregulował swoją maszynę, żeby udzielała właściwych odpowiedzi; nie musi wcale znać szczegółowej metody działania swojego algorytmu.

To kolejny algorytm uczenia maszynowego, podobny do lasu losowego, który poznaliśmy w rozdziale *Wymiar sprawiedliwości*. Maszyna taka

wykracza poza to, do czego zaprogramował ją jej operator, i sama się uczy z pokazywanych sobie obrazów. Właśnie ta umiejętność uczenia się nadaje naszemu algorytmowi aurę „sztucznej inteligencji". A że liczne warstwy pokręteł i suwaków tworzą z sieci strukturę głęboką – stąd określenie „głębokie uczenie" (*deep learning*).

Sieci neuronowe istnieją od połowy XX wieku, lecz jeszcze do niedawna nie mieliśmy łatwego dostępu do naprawdę mocnych komputerów, niezbędnych do tego, żeby wycisnąć z nich wszystko, co się da. Świat jednak musiał się w końcu przebudzić i potraktować je poważnie w 2012 roku, kiedy informatyk Geoffrey Hinton z dwójką swoich studentów zgłosił do konkursu rozpoznawania obrazów nowy rodzaj sieci neuronowej[12]. Zadanie zgłoszonych do konkursu maszyn polegało na rozpoznawaniu między innymi psów. Sztucznie inteligentny algorytm Hintona i jego studentów rozłożył na łopatki najlepszych rywali i zapoczątkował wielki renesans w uczeniu głębokim.

Algorytm, o którego sposobie podejmowania decyzji nie mamy zielonego pojęcia, można by oskarżyć o czary, a przecież sami uczymy się bardzo podobnie. Rozważmy następujące porównanie. Zespół badaczy wytrenował niedawno algorytm do odróżniania zdjęć wilków i małych psów husky. Badacze zademonstrowali, iż algorytm tak ponastawiał swoje wewnętrzne mechanizmy, że nie korzystał z żadnych informacji mających jakikolwiek związek z psami – decyzje opierał wyłącznie na tym, czy na drugim planie zdjęcia widnieje śnieg. Śnieg – wilk. Brak śniegu – husky[13].

Wkrótce po opublikowaniu pracy wspomnianego zespołu rozmawiałam z Frankiem Kellym, profesorem matematyki na uniwersytecie w Cambridge, który opowiedział mi o rozmowie, jaką odbył niedawno ze swoim czteroletnim wnukiem. W drodze do przedszkola spotkali psa husky. Chłopiec powiedział wtedy, że pies „wygląda jak wilk". Kiedy Frank spytał go, skąd wie, że to nie wilk, wnuczek odpowiedział: „Bo jest na smyczy".

Czuły sojusz

Po dobrym algorytmie rozpoznawania raka piersi spodziewasz się dwóch rzeczy: ma być wystarczająco c z u ł y, żeby wychwycić wszystkie anomalie

we wszystkich piersiach, które mają guzy; nie może ślizgać się po pikselach i oznajmiać, że skan jest czysty. Ale ma być również na tyle s w o i s t y, żeby nie oznaczał całkowicie normalnych tkanek piersiowych jako podejrzanych.

Poznaliśmy już zasady czułości i swoistości w rozdziale *Wymiar sprawiedliwości*. To bliscy kuzyni wyniku fałszywie pozytywnego i wyniku fałszywie negatywnego – albo Dartha Vadera i Luke'a Skywalkera, jak moim zdaniem powinno się je nazywać w literaturze naukowej. W omawianym tu kontekście o wyniku fałszywie pozytywnym mówimy zawsze, kiedy zdrowa kobieta zostaje powiadomiona, że ma raka piersi, a o wyniku fałszywie negatywnym – kiedy kobieta z guzem dostaje informację, że jest zupełnie zdrowa. Test swoisty nie będzie dawał prawie żadnych wyników fałszywie pozytywnych, podczas gdy test czuły będzie miał bardzo znikomy odsetek wyników fałszywie negatywnych. Nie ma znaczenia, do czego zaprzągnęłaś swój algorytm: może prognozować ryzyko recydywy, diagnozować raka piersi czy też (jak zobaczymy w rozdziale *Przestępczość*) rozpoznawać prawidłowości w zachowaniu przestępców – ale historia zawsze wygląda tak samo. Dążymy do tego, żeby mieć jak najmniej wyników fałszywych – i pozytywnych, i negatywnych.

Kłopot w tym, że wyregulowanie algorytmu często oznacza wybór między czułością a swoistością. Skupiając się na poprawie funkcjonowania w pierwszym zakresie, możesz pogorszyć jakość w drugim. Jeżeli na przykład zdecydujesz się postawić na całkowite wyeliminowanie wyników fałszywie negatywnych, twój algorytm może zacząć uznawać za podejrzane wszystkie skany piersi, które mu pokażesz. Osiągniesz stuprocentową czułość, co z pewnością uznasz za satysfakcjonujący wynik, będzie to jednak oznaczać, że ogromna liczba zdrowych kobiet zostanie niepotrzebnie poddana leczeniu. Powiedzmy zatem, że zdecydujesz się postawić na całkowite wyeliminowanie wyników fałszywie pozytywnych – wówczas twój algorytm przepuści przez sito wszystkie skany jako obrazy zdrowych piersi kobiecych, uzyskując stuprocentową swoistość. Cudownie! Chyba że jesteś jedną z tych kobiet z guzem w piersi, który algorytm właśnie zignorował.

Ciekawe, że patologowie nie mają zazwyczaj problemów ze swoistością. Prawie nigdy nie uznają zdrowych komórek za rakowe. Trzeba natomiast

przyznać, że my, ludzie, mamy pewne kłopoty z czułością. Niepokojąco często zdarza się nam przeoczyć małego guza, nawet wyraźnie złośliwego.

Nasze ludzkie słabości zostały mocno uwypuklone w przeprowadzonym niedawno pojedynku człowieka z algorytmem. Zespoły komputerów z całego świata stanęły do rywalizacji z patologiem. Uczestnicy turnieju CAMELYON16 mieli odnaleźć wszystkie guzy kryjące się w czterystu skrawkach biopsyjnych. Dla ułatwienia wszystkie przypadki należały do obu skrajności: tkanka normalna albo inwazyjny rak piersi. Patolog nie miał żadnego limitu czasu: mógł analizować każdą próbkę tak długo, jak uznał to za stosowne. Człowiek, zgodnie z oczekiwaniami, w sumie poprawnie wywiązał się z zadania (96 procent trafności)[14], i to bez wyników fałszywie pozytywnych. Przepuścił jednak bardzo wiele maleńkich komórek rakowych ukrytych w tkance piersiowej, rozpoznając zaledwie 73 procent z nich w ciągu 30 godzin analizy.

Sednem problemu nie była horrendalna liczba malutkich pikseli do sprawdzenia – człowiek potrafi przeoczyć oczywistą anomalię, nawet kiedy ma ją tuż pod nosem. W 2013 roku badacze z Harvardu ukryli na jednym z kilkuset radiogramów klatki piersiowej podobiznę goryla, po czym poprosili grupę 24 nieświadomych wybiegu radiologów o sprawdzenie, czy na zdjęciach nie widać śladów raka. Osiemdziesiąt trzy procent badanych nie dostrzegło goryla, choć analiza ruchów gałek ocznych wykazała, że większość z nich musiała na niego patrzeć[15]. Zresztą sprawdźcie sami na poniższym zdjęciu[16].

Z algorytmami jest akurat odwrotnie: są wręcz nadgorliwe w wykrywaniu anormalnych skupisk komórek, nawet tych całkowicie zdrowych. Najlepsza zgłoszona do turnieju CAMELYON16 sieć neuronowa zdołała osiągnąć imponujący wynik 92,4 procent wykrytych guzów[17], dając jednak przy tym aż osiem wyników fałszywie pozytywnych na skrawek, czyli niepoprawnie uznając skupiska normalnych komórek za podejrzane. Przy tak słabej swoistości najnowocześniejsze algorytmy zdecydowanie skłaniają się ku strategii diagnostycznej pod tytułem „każda kobieta ma raka piersi!" i po prostu nie są jeszcze gotowe do pisania samodzielnych raportów histopatologicznych.

Jest jednak dobra wiadomość: nikt nie będzie ich o to prosić. Zależy nam jedynie na połączeniu mocnych stron człowieka i maszyny. Algorytm ma wykonywać mrówczą robotę, przerabiając niewiarygodne ilości

zakodowanych w próbkach informacji i zaznaczając zmiany w kilku kluczowych dla nas obszarach zainteresowania. Wtedy pałeczkę przejmie patolog. To bez znaczenia, że maszyna zaznaczy także komórki nierakowe; człowiek bardzo szybko wyeliminuje te błędy. Tego rodzaju algorytmiczna pomoc we wstępnej obróbce materiału nie tylko oszczędzi nam mnóstwo czasu, lecz także wywinduje nam ogólną trafność na astronomiczny poziom 99,5 procent[18].

Brzmi to pewnie rewelacyjnie, ale faktem jest, że ludzie zawsze byli dobrzy w diagnozowaniu nowotworów agresywnych. Twardszy orzech do zgryzienia stanowiły nieoczywiste przypadki ze środka spektrum, gdzie rozróżnienie między rakiem i nierakiem jest bardziej subtelne. Czy algorytmy i tutaj nam pomogą? Odpowiedź brzmi: (prawdopodobnie) tak. Pod warunkiem że algorytm nie będzie próbował diagnozować, posiłkując się naszymi ludzkimi, niepewnymi kategoriami. Niewykluczone, że maszyna, która bije nas na głowę w wykrywaniu anomalii maleńkich wycinków danych, mogłaby podsunąć nam ogólnie lepszą metodę diagnozowania. Robiąc rzeczy, których lekarze nie potrafią.

Siostry pod lupą

W 1986 roku David Snowdon, epidemiolog z University of Kentucky, zdołał nakłonić 678 sióstr zakonnych, członkiń Zgromadzenia Sióstr Szkolnych de Notre Dame, do oddania mu swoich mózgów. Chodziło o udział w niezwykłym badaniu przyczyn choroby Alzheimera.

Każda z kobiet, które na początku badania liczyły od 75 do 103 lat, zgodziła się raz w roku aż do końca życia poddawać się kilku testom pamięciowym i niezależnie od tego, czy na początku projektu miała objawy demencji, czy też nie, wyraziła zgodę na to, żeby po jej śmierci zespół Snowdona pobrał jej najcenniejszy narząd i przebadał go pod kątem oznak choroby[19].

Dzięki wspaniałomyślności sióstr powstała niezwykła baza danych. Ponieważ żadna z kobiet nie miała dzieci, nie paliła i w zasadzie nie piła, naukowcy mogli wyeliminować wiele czynników zewnętrznych uznawanych powszechnie za zwiększające prawdopodobieństwo zachorowania na alzheimera. A ponieważ wszystkie prowadziły podobny

tryb życia, miały podobny dostęp do opieki zdrowotnej i cieszyły się podobnym wsparciem społecznym, stanowiły dla badaczy doskonały materiał porównawczy.

Kilka lat po rozpoczęciu badania członkowie zespołu Snowdona stwierdzili, że ta wyjątkowa grupa eksperymentalna pozwala im się podłączyć do jeszcze innego zbioru bezcennych danych. Wiele z tych zaawansowanych już wiekiem sióstr musiało w młodości, jeszcze przed złożeniem ślubów, przedłożyć zakonowi odręcznie napisany życiorys. Po wyliczeniu średniej okazało się, że kobiety pisały go, mając 22 lata, czyli kilkadziesiąt lat przed wystąpieniem pierwszych objawów demencji. I choć brzmi to niewiarygodnie, naukowcy odkryli w życiorysach sióstr wskazówki pozwalające już wówczas przewidzieć, co się wydarzy w odległej przyszłości.

Zespół Snowdona przeanalizował każdy życiorys pod kątem złożoności języka i stwierdził korelację między umiejętnością wysławiania się kobiet w młodości a prawdopodobieństwem ich zachorowania na demencję w podeszłym wieku.

Oto przykładowe zdanie z życiorysu zakonnicy, która zachowała wysoką sprawność umysłową do końca życia: „Po skończeniu ósmej klasy w 1921 roku zapragnęłam zostać aspirantką w Mankato, sama jednak nie miałam odwagi, aby spytać rodziców o zgodę, więc siostra Agreda zrobiła to w moim imieniu i moi rodzice skwapliwie na to przystali".

Porównaj to ze zdaniem napisanym przez siostrę, która pod koniec życia coraz słabiej wypadała w dorocznych testach pamięciowych: „Po skończeniu szkoły pracowałam na poczcie".

Korelacja była tak wyraźna, że badacze mogli przewidywać demencję wyłącznie na podstawie tych życiorysów. Dziewięćdziesiąt procent sióstr, które zachorowały na alzheimera, wykazywało w młodości „słabą sprawność językową", podczas gdy „niska gęstość idei"[20] cechowała prace zaledwie 13 procent zakonnic, które zachowały wysoką sprawność umysłową do późnej starości.

Studium to pokazuje również, jak niewiarygodnie długą drogę mamy jeszcze do przebycia, żeby poznać ludzkie ciało. Samo odkrycie korelacji sprawności umysłowej ze skłonnościami do alzheimera nie daje nam

jeszcze odpowiedzi na pytanie, d l a c z e g o wysoka sprawność umysłowa zapobiega chorobie. (Czy to dobre wykształcenie chroni nas przed demencją? Czy może osoby z większymi skłonnościami do alzheimera po prostu wolą wysławiać się prościej?) Niemniej odkrycia zespołu Snowdona sugerują, że rozwój choroby może być rozłożony na kilkadziesiąt lat.

Dla nas ważniejsze jest jednak stwierdzenie, że subtelne sygnały o przyszłym stanie naszego zdrowia mogą być zakodowane w bardzo małych porcjach danych, z pozoru niezwiązanych z obszarem naszych zainteresowań – i to na wiele lat przed wystąpieniem pierwszych objawów choroby. Możemy tylko zgadywać, jak wielkie możliwości będą mieć medyczne algorytmy przyszłości potrafiące pracować na gigantycznych bazach danych. Nie możemy wykluczyć, że pewnego dnia będą umiały dostrzec pierwsze ślady raka na wiele lat przed nami.

Sztuka przewidywania

Pod koniec lat siedemdziesiątych ubiegłego stulecia trójka duńskich patologów z miasta Ribe wykonała obustronną mastektomię na grupie zwłok 83 kobiet w wieku od 22 do 89 lat; 6 z nich zmarło z powodu raka inwazyjnego piersi. Nic dziwnego, że gdy patologowie preparowali wycinki do badań – dzieląc je na cztery części, a potem krojąc każdą z nich na cienkie skrawki – próbki pobrane od tych sześciu zmarłych na raka kobiet nosiły wszelkie znamiona choroby. Duńczycy odkryli jednak ze zdumieniem, że u niemal czwartej części pozostałych kobiet – które zmarły z zupełnie innych przyczyn, na przykład na choroby serca lub w następstwie wypadku samochodowego – wykryto takie same sygnały ostrzegające przed rakiem piersi, jakich patologowie szukają u żywych pacjentek.

U 14 kobiet, które nie miały za życia żadnych objawów choroby, odkryto komórki rakowe *in situ*, które nie wyszły poza przewody mleczne i gruczoły i które uznano by za nowotwory złośliwe, gdyby kobiety zdążyły poddać się badaniu piersi. U trzech kobiet stwierdzono w tkance piersiowej komórki atypowe, które podczas badania patologicznego również uznano by za zagrożenie dla zdrowia, a jedna kobieta miała raka inwazyjnego piersi, choć do śmierci o tym nie wiedziała[21].

Dane były zaskakujące, ale wyniki nie były przypadkowe – inni badacze uzyskali podobne rezultaty. Okazuje się, że mniej więcej 9 procent kobiet może nie wiedzieć o tym, że nosi w piersi guza[22] – a to mniej więcej dziesięciokrotność odsetka kobiet z faktycznie zdiagnozowanym rakiem piersi[23].

Co się zatem dzieje? Czy mamy do czynienia z cichą epidemią? Doktor Jonathan Kanevsky, innowator medyczny i rezydent kliniki chirurgii Uniwersytetu McGill w Montrealu, zaprzecza. A przynajmniej nie do końca się z tym zgadza. Obecność w organizmie komórek rakowych nie musi od razu stanowić zagrożenia: „Jeżeli ktoś ma komórkę rakową, są duże szanse, że jego układ odpornościowy rozpozna ją jako mutację, po czym zaatakuje i zniszczy – i ten rak nigdy nie przerodzi się w nic groźnego. Czasem jednak układ odpornościowy tak się rozreguluje, że organizm wręcz wspiera rozwój raka. W takiej sytuacji rak może rzeczywiście zabić”[24].

Guz guzowi nierówny. Jedne zostaną zneutralizowane przez organizm, drugie będą sobie siedzieć cichutko gdzieś w tkance aż do naszej śmierci, jeszcze inne rozwiną się w agresywną postać nowotworu. Kłopot w tym, że najczęściej mamy bardzo skąpe przesłanki do stwierdzenia, który jest którym.

I właśnie dlatego nieoczywiste kategorie zajmujące całą przestrzeń między biegunami łagodnej i najbardziej złośliwej postaci nowotworu stanowią dla nas tak wielkie wyzwanie. To jedyna klasyfikacja, z jaką mają do czynienia lekarze, ale gdy patolog odkryje w badanej próbce grupę nieco podejrzanych komórek, etykieta, na którą się zdecyduje, pomoże jedynie opisać obecny stan analizowanej tkanki piersiowej. To niezbyt wielka pomoc, bo człowiek wolałby raczej dowiedzieć się czegoś więcej o swojej najbliższej przyszłości. A to właśnie ona najbardziej martwi pacjentki z podejrzeniem raka piersi.

Rezultat jest taki, że ludzie są często przesadnie ostrożni w wybranej przez siebie metodzie leczenia. Weźmy na przykład raka *in situ*. To kategoria znajdująca się bliżej tego bardziej niepokojącego bieguna spektrum, kiedy nowotwór jest obecny, lecz nie ma jeszcze nacieków na otaczającą przewody tkankę. Choć brzmi to poważnie, okazuje się, że tylko

mniej więcej jeden na dziesięć nowotworów *in situ* przekształci się w guza, który mógłby zabić. Mimo to jedna czwarta kobiet, które słyszą taką diagnozę w Stanach Zjednoczonych, przechodzi pełną mastektomię – a to poważna operacja, która dramatycznie zmienia ich życie nie tylko fizycznie, lecz często również psychicznie[25].

Nie ulega wątpliwości, że im bardziej agresywnie będzie się prowadzić badania przesiewowe piersi, tym więcej wyłapie się kobiet, które bez tych badań żyłyby sobie spokojnie bez większych zmartwień, w błogiej nieświadomości, że mają w piersi nieszkodliwego guza. Pewien niezależny panel brytyjski doszedł do wniosku, że na każde 10 tysięcy kobiet, które zostaną zaproszone na badania mammograficzne w ciągu najbliższych 20 lat, 43 uda się uratować przed śmiercią z powodu raka. Studium opublikowane w „New England Journal of Medicine" zakończyło się konkluzją, że na każde 100 tysięcy kobiet, które zgłaszają się na badania mammograficzne, potencjalnie niebezpieczne dla życia guzy wykrywa się u 30 pań[26]. Ale aż od trzech do czterech razy więcej kobiet padnie ofiarą nadrozpoznania i będzie leczonych z powodu guza, który nigdy nie stanowiłby dla nich śmiertelnego zagrożenia[27].

Problem nadrozpoznania i zbędnego leczenia trudno rozwiązać, kiedy dobrze nam idzie wykrywanie anomalii, lecz znacznie gorzej radzimy sobie z prognozowaniem, w którą stronę te anomalie się rozwiną. Nie traćmy jednak nadziei. Być może – podobnie jak to było z życiorysami zakonnic – w informacjach z przeszłości i teraźniejszości pacjenta zdołamy kiedyś odnaleźć ciche podpowiedzi o stanie jego zdrowia za wiele, wiele lat. A jeśli tak, to wyłuskanie takich informacji z bazy danych będzie idealną robotą dla sieci neuronowej.

Na polu medycyny, gdzie lekarze od dziesięcioleci zmagają się z pytaniem, dlaczego jedna anomalia jest groźniejsza od innej, algorytm, któremu nie kazano szukać czegoś konkretnego, może nam oddać nieocenione usługi. Wystarczy, że zgromadzi się wystarczająco duży zbiór próbek histopatologicznych – w tym pewną liczbę skrawków guzów, które spowodowały metastazę, czyli przerzuty do innych części ciała, oraz takich, które przerzutów nie miały – i każe się sieci neuronowej szukać ukrytych wskazówek do przyszłego stanu zdrowia pacjenta, nie ograniczając jej żadnymi uprzedzeniami towarzyszącymi każdej teorii.

Zdaniem Jonathana Kanevskiego: „To algorytm ustalałby niepowtarzalne cechy każdego skanu, które decydują o tym, czy guz będzie metastatyczny, czy nie”[28].

Dla takiego algorytmu kategoryzowanie każdej próbki miałoby dużo mniejsze znaczenie niż obecnie. Nie musielibyśmy zawracać sobie głowy pytaniami „dlaczego?” i „jak?” – moglibyśmy od razu przejść do informacji, która interesuje nas najbardziej: czy poddać się leczeniu, czy nie?

Mam dobrą wiadomość: prace nad stworzeniem takiego algorytmu już się rozpoczęły. Andy Beck, wspomniany już wcześniej patolog z Harvardu i dyrektor generalny firmy PathAI, niedawno „rozluźnił” nieco swój algorytm podczas analizy serii próbek pacjentów z Holandii i stwierdził, że najlepsze prognostyki przeżycia dawała analiza nie samego guza, lecz anomalii otaczającej go tkanki[29]. To ważne odkrycie i dobry przykład na to, że algorytmy same mogą przecierać nowe szlaki w badaniach nad rakiem i odnajdywać prawidłowości, dzięki którym możemy doskonalić sztukę przewidywania.

Poza tym dysponujemy dziś nieprzebranym bogactwem danych, z których możemy czerpać. Dzięki prowadzonym na całym świecie programom przesiewowych badań mammograficznych mamy prawdopodobnie więcej skanów tkanki piersiowej niż któregokolwiek innego organu ludzkiego ciała. Nie jestem patologiem, ale każdy ekspert, z którym rozmawiałam, przekonywał mnie, że umiejętność trafnego przewidywania, czy podejrzana próbka przekształci się w raka, to już tylko kwestia czasu. To bardzo, bardzo możliwe, że zanim ta książka zostanie wydana, ktoś gdzieś urzeczywistni tę ideę i zmieni nasz świat.

Cyfrowa diagnoza

Idee te wykraczają daleko poza kwestię raka piersi. Sieci neuronowe, które budują Andy Beck i inni, nie są wybredne – można by podsunąć im i kazać skategoryzować cokolwiek: psy, kapelusze lub sery. Trzeba tylko informować maszynę, kiedy poprawnie, a kiedy błędnie oceniła treść zdjęcia, aż w końcu się nauczy. A skoro ta rodzina algorytmów radzi sobie tak dobrze, że można już myśleć o ich stosowaniu w praktyce, będą one miały wpływ na wiele innych gałęzi współczesnej medycyny.

Oto przykład. Wielki sukces odniósł niedawno zespół Google Brain, który skonstruował algorytm wykrywający retinopatię cukrzycową – chorobę niszczącą naczynia krwionośne w warstwie światłoczułej oka i stanowiącą dziś w skali globalnej najczęstszą przyczynę ślepoty, której można zapobiegać. Jeżeli pacjent wie, że cierpi na retinopatię cukrzycową, może serią zastrzyków uratować wzrok, jeśli jednak choroba nie zostanie w porę wykryta, może prowadzić do nieodwracalnej ślepoty. W Indiach, gdzie dostęp do specjalistów mogących ją zdiagnozować jest bardzo ograniczony, 45 procent chorych utraci część widzenia, zanim się dowie, że choruje. Algorytm zespołu Google'a, który powstał we współpracy z lekarzami indyjskimi, stawia obecnie równie trafne diagnozy jak okulista.

Podobne algorytmy diagnozują już choroby serca[30], przewlekłą obturacyjną chorobę płuc[31], udar mózgu[32] i czerniaka[33]. Istnieją nawet systemy, które podczas zabiegu kolonoskopii na bieżąco diagnozują polipy.

To już pewne: jeżeli można chorobę sfotografować i fotografię opisać, można też stworzyć algorytm rozpoznający ją na zdjęciach. Bardzo prawdopodobne, że taki algorytm będzie stawiał dokładniejsze (i zapewne wcześniejsze) diagnozy niż człowiek w białym kitlu.

A co z bardziej chaotycznymi formami danych medycznych? Czy sukcesy algorytmów da się przenieść na inne pola, poza bardzo wyspecjalizowane i wąsko zakreślone zadania? Czy maszyna mogłaby szukać znaczących wskazówek na przykład w notatkach naszej lekarki? Albo w tym, jak opisujemy swój ból?

Co powiecie na absolutne medyczne science fiction, w którym maszyna w waszej przychodni uważnie wysłuchuje opisu waszych objawów i analizuje waszą historię choroby? Co powiecie na maszynę, która czuje się jak ryba w wodzie na froncie najbardziej zaawansowanych badań medycznych? Która stawia dokładną diagnozę i podaje idealnie dopasowany do pacjenta plan leczenia?

Krótko mówiąc: co powiecie na coś w rodzaju Watsona firmy IBM?

To elementarne, drogi Watsonie

W 2004 roku Charles Lickel z zapałem zabierał się do steka w towarzystwie kolegów w nowojorskiej restauracji. W trakcie kolacji sala zaczęła nieoczekiwanie pustoszeć. Zaintrygowany Charles poszedł za resztą klientów, którzy ciasnym kołem otoczyli telewizor, gdzie zaczynał się właśnie popularny teleturniej *Jeopardy!*[11*]. Niekwestionowany mistrz gry Ken Jennings miał szansę przedłużyć rekordową, sześciomiesięczną passę, więc wszyscy chcieli popatrzeć, jak mu pójdzie[34].

Charles Lickel był wiceprezesem działu oprogramowania firmy IBM. Od kilku lat, odkąd Deep Blue pokonał Garry'ego Kasparowa w szachach, szefowie wiercili mu dziurę w brzuchu, żeby znalazł koncernowi nowe wyzwanie, godne jego pozycji. Stojąc w restauracji i patrząc na ludzi zafascynowanych zmaganiami mistrza teleturnieju, Lickel zaczął się zastanawiać, czy dałoby się stworzyć maszynę zdolną pokonać takiego mistrza.

Nie było to wcale zadanie łatwe. Stworzenie Watsona zajęło siedem lat, ale udało się. Watson miał stanąć do pojedynku z Kenem Jenningsem w specjalnym wydaniu teleturnieju. Maszyna pokonała człowieka. W trakcie projektu „Watson" firma IBM weszła na drogę prowadzącą do stworzenia pierwszej na świecie uniwersalnej maszyny diagnostycznej. Wrócimy do niej za chwilę. Tymczasem pozwolę sobie omówić kilka kluczowych idei, które przyświecały twórcom Watsona i zapoczątkowały trend konstruowania medycznych algorytmów diagnostycznych.

Kto nigdy nie słyszał o *Jeopardy!*, musi wiedzieć, że to popularny amerykański teleturniej w formie odwróconego quizu z wiedzy ogólnej: zawodnicy dostają wskazówki w formie odpowiedzi, do których muszą sformułować pytania. Na przykład w kategorii „słowa wieloznaczne" wskazówka mogłaby brzmieć następująco:

> Rozpatrywać jakąś sprawę, zastanawiać się nad czymś, przemyśliwać coś; także dzielić na ściśle określone porcje za pomocą specjalnego urządzenia.

Żeby sformułować właściwe pytanie – „Co znaczy rozważać?" – algorytmiczny gracz musiałby się nauczyć przedzierać przez kilka warstw odpowiedzi. Po pierwsze, na tyle dobrze rozumieć język, żeby szukać sensu w pytaniu-odpowiedzi i skojarzyć, że „rozpatrywanie", „zastanawianie się", „przemyśliwanie" i „dzielenie za pomocą specjalnego urządzenia" to odrębne elementy wskazówki. Już to stanowi dla algorytmu ogromne wyzwanie.

A to był dopiero pierwszy krok. Następnie Watson musiał poszukać potencjalnych kandydatów pasujących do każdego z elementów pytania-odpowiedzi. Element „rozpatrywać" mógłby dać całą gamę potencjalnych odpowiedzi: „brać pod uwagę", „rozważać", „selekcjonować", ale także, na przykład, „oceniać". Watson musiał – nomen omen – rozważyć po kolei każdą możliwość i sprawdzić, jak pasuje do wszystkich pozostałych wskazówek. Raczej nie znalazł potwierdzenia na to, że „zastanawiać się" można skojarzyć z „dzieleniem na ściśle określone porcje", ale już słowo „rozważać" zdecydowanie tak, co pozwoliło mu z większą pewnością upatrywać w tym właśnie słowie właściwą odpowiedź. Na końcu, po połączeniu tych wszystkich informacji, Watson musiał postawić wszystko na jedną kartę i wybrać jedno słowo.

Choć udział w teleturnieju wiedzy ogólnej jest wyzwaniem raczej trywialnym w porównaniu z diagnozowaniem chorób, wymaga jednak zastosowania identycznych mechanizmów logicznych. Wyobraźmy sobie, że idziemy do lekarza, bo straciliśmy na wadze, chociaż nie jesteśmy na diecie. Boli nas brzuch i na dodatek jeszcze męczy nas czasem zgaga. Analogicznie jak w przypadku gry w teleturnieju, wyzwanie będzie polegać na znalezieniu potencjalnych diagnoz (odpowiedzi), które mogą tłumaczyć objawy (wskazówki), a następnie na wyszukaniu potwierdzenia oraz zaktualizowaniu stopnia pewności dla każdej hipotezy. Lekarze nazywają to diagnostyką różnicową, matematycy zaś – wnioskowaniem bayesowskim[12*].

Narodziny Watsona Mistrza Quizów wcale nie gwarantowały, że stworzenie Watsona Geniusza Medycyny będzie zadaniem prostym. Niemniej ogłaszając publicznie zamiar wejścia w służbę zdrowia, IBM nie zawahał się użyć wielkich słów. Cały świat się dowiedział, że ostatecznym celem Watsona jest „wyeliminowanie raka"[35], a wynajęty przez firmę

słynny aktor Jon Hamm zachwalał go jako „jedno z najpotężniejszych narzędzi stworzonych przez człowieka".

Co jak co, ale taka wizja medycznej utopii powinna być inspiracją dla nas wszystkich. Z tym że w przypadku Watsona cała para poszła w gwizdek.

Pierwsza prestiżowa umowa z centrum walki z rakiem MD Anderson Cancer Center przy University of Texas została wypowiedziana w 2016 roku. Rozeszły się pogłoski, że mimo zainwestowania w projekt 62 milionów dolarów[36] i czterech lat pracy Watson nadawał się jedynie do ściśle nadzorowanych testów pilotażowych. Jakiś czas później, na początku września 2017 roku, śledztwo przeprowadzone przez internetowy portal zdrowia STAT wykazało, że Watson „nadal nie potrafi zrobić pierwszego kroku, jakim jest rozpoznanie różnych form raka"[37].

Przykra sprawa.

Trzeba jednak uczciwie przyznać, że nie było aż tak tragicznie. W Japonii Watson zdołał zdiagnozować u pewnej kobiety rzadką formę białaczki, której nie dostrzegli lekarze[38], a jego analiza doprowadziła do odkrycia pięciu genów powiązanych ze stwardnieniem zanikowym bocznym[39]. Generalnie jednak programiści IBM nie zdołali spełnić obietnic swoich nadpobudliwych kolegów z działu marketingu.

Trudno nie kibicować firmie, która mierzy siły na takie zamiary. Teoretycznie zbudowanie maszyny diagnozującej choroby (a nawet sugerującej pacjentom rozsądne plany leczenia) jest możliwe – i jeśli komuś przyświeca taki szczytny cel, można go za to tylko podziwiać. Z drugiej strony to naprawdę szalenie trudne zadanie. Znacznie trudniejsze niż wygranie teleturnieju i znacznie, znacznie trudniejsze od rozpoznawania komórek rakowych na skanie.

Mogłoby się wydawać, że aby stworzyć wszechstronną maszynę diagnostyczną, wystarczy pójść krok dalej w logice algorytmów wykrywających raka na zdjęciach. Algorytmy te mają jednak wielką przewagę: dostają do zbadania rzeczywiście istniejące komórki, które mogą spowodować problemy. Maszyna diagnostyczna natomiast otrzymuje tylko rozproszone informacje, nieraz odległe od pierwotnej przyczyny problemu. Pacjentka może odczuwać mrowienie spowodowane

skurczem mięśni spowodowanym uciskiem na nerw spowodowanym dźwiganiem zbyt ciężkich przedmiotów. Albo pacjent ma krew w kale z powodu hemoroidów, które pojawiły się wskutek zaparć spowodowanych niewłaściwą dietą. Algorytm (albo lekarz) musi skupić się na jednym objawie i prześledzić cały łańcuch przyczynowo-skutkowy prowadzący do poprawnej diagnozy. To właśnie musi zrobić Watson – dla maszyny jest to zadanie niewyobrażalnie trudne.

A to jeszcze nie koniec trudności.

Pamiętacie sieć neuronową rozpoznającą psy husky i wilki? Wytrenowanie jej do wykonywania tego zadania było łatwe. Cała praca programistów sprowadzała się do znalezienia stosu zdjęć, opatrzenia ich etykietami „husky" albo „wilk" i załadowania ich do maszyny. Zbiór danych był prosty i niedwuznaczny. „Ale w przypadku wyspecjalizowanej gałęzi medycyny – powiedział w wywiadzie dla «MIT Technology Review» Thomas Fuchs, specjalista od patologii cyfrowej – do poprawnego opisania informacji ładowanych do komputera czasem potrzeba ekspertów z dwudziestoletnim stażem"[40].

Nie jest to przeszkoda nie do pokonania, jeżeli bardzo ściśle ująć problem (na przykład dzieląc skrawki histopatologiczne na dwie grupy: „bezsprzecznie łagodne" i „potwornie złośliwe"). Ale wszystkowidząca maszyna diagnostyczna pokroju Watsona musiałaby najpierw zrozumieć praktycznie każdą możliwą chorobę. Wymagałoby to zatrudnienia prawdziwej armii nadzwyczaj wysoko wykwalifikowanych specjalistów, gotowych przez bardzo długi czas wprowadzać do komputera informacje o pacjentach i szczegółach ich dolegliwości. A tacy specjaliści mają na ogół pełne ręce roboty.

I dopiero teraz dochodzimy do końcowego problemu. Najtrudniejszego ze wszystkich.

Kłopoty z danymi

Tamara Mills była jeszcze bardzo mała, kiedy jej rodzice po raz pierwszy zauważyli, że coś jest nie tak z jej oddychaniem. Kiedy dziewczynka miała dziewięć miesięcy, lekarze już rozpoznali u niej astmę – chorobę, która dotyka 5,4 miliona osób w Wielkiej Brytanii i 25 milionów w Stanach

Zjednoczonych[41]. Mimo że Tamara była młodsza niż większość astmatyków, objawy choroby na tak wczesnym etapie dało się łatwo kontrolować i podobnie jak wiele innych dzieci chorych na astmę, Tamara spędziła dzieciństwo na zabawach nad brzegiem morza w północnej Anglii (choć zawsze musiała mieć pod ręką swój inhalator).

Kiedy Tamara miała osiem lat, zaraziła się świńską grypą, którą przeszła bardzo źle. Później się okazało, że był to punkt zwrotny w jej życiu. Od tej pory jedna infekcja dróg oddechowych goniła następną. Niekiedy podczas ataków astmy usta Tamary robiły się sine. I choć dziewczynka raz za razem wracała z matką do lekarza i miejscowego szpitala, choć rodzice dziewczynki bez przerwy narzekali, że Tamara zużywa cały zapas inhalatorów szybciej, niż można je przepisywać[42], żaden z lekarzy nie skierował jej nigdy do specjalisty.

Rodzina i nauczyciele Tamary mieli jednak świadomość, że sprawa robi się poważna. Po dwóch bardzo groźnych atakach, po których dziewczynka wylądowała w szpitalu, została zwolniona z lekcji wychowania fizycznego. Kiedy schody w domu stały się dla niej przeszkodą nie do pokonania, zamieszkała z dziadkami w parterowym domu.

Dziesiątego kwietnia 2014 roku Tamara zapadła na kolejne zapalenie dróg oddechowych. W nocy dziadek usłyszał, że wnuczka nie może złapać oddechu. Wezwał karetkę i starał się ją ratować za pomocą dwóch inhalatorów i zbiornika z tlenem, lecz stan dziewczynki ciągle się pogarszał. Tamara zmarła tej samej nocy, mając zaledwie 13 lat.

Astma nie jest chorobą śmiertelną, a mimo to każdego roku umiera na nią w Wielkiej Brytanii 1200 osób, z czego 26 to dzieci[43]. Ocenia się, że w dwóch trzecich z tych przypadków można było zapobiec śmierci. Dotyczy to również Tamary. Wszystko zależy jednak całkowicie od wczesnego rozpoznania sygnałów ostrzegawczych i podjęcia stosownych działań.

W ciągu czterech lat poprzedzających tamten ostatni, śmiertelny atak Tamara była u lokalnych lekarzy i w lokalnym szpitalu ni mniej, ni więcej, tylko 47 razy. Było oczywiste, że plan jej leczenia nie skutkuje, a jednak za każdym razem, kiedy zwracała się o pomoc medyczną, pracownicy służby zdrowia zajmowali się wyłącznie ostrymi objawami jej choroby. Nikt nie

spojrzał na sprawę z innej perspektywy. Nikt nie dostrzegł prawidłowości w coraz częstszych wizytach; nikt nie zauważył, że stan jej zdrowia bezustannie się pogarsza; nikt nie wystąpił z propozycją, żeby spróbować czegoś nowego[44].

Wiemy też, dlaczego tak się działo. Możecie wierzyć lub nie (choć każdy, kto mieszka w Wielkiej Brytanii, z pewnością uwierzy), ale w brytyjskiej służbie zdrowia nie kompletuje się dokumentacji medycznej w ramach standardowych procedur. Jeżeli wylądujesz w szpitalu, pracujący tam specjaliści nie będą wiedzieli nic o twoich wcześniejszych wizytach u lekarza ogólnego. Znaczna część dokumentacji nadal jest papierowa, co oznacza, że sposób jej udostępniania nie zmienił się od dziesięcioleci. To jedna z przyczyn, dla których brytyjska służba zdrowia dzierży niechlubny rekord świata w liczbie nabywanych telefaksów[45].

Brzmi to pewnie niewiarygodnie, lecz nie jesteśmy sami. W Stanach Zjednoczonych jest mnóstwo prywatnych gabinetów i wielkich sieci szpitali, które nie są ze sobą połączone; inne kraje, takie jak Niemcy, zaczęły już tworzyć elektroniczne kartoteki pacjentów, mamy jednak przed sobą jeszcze długą drogę, zanim stanie się to światową normą. W przypadku Tamary Mills brak historii choroby oznaczał, że żaden lekarz nie zdawał sobie w pełni sprawy ze stopnia zaawansowania jej astmy. Ewentualne rozwiązanie tego poważnego błędu systemowego przyjdzie już, niestety, za późno dla Tamary. Niemniej scalanie danych pacjentów pozostaje wielkim wyzwaniem. Maszyna pokroju Watsona mogłaby uratować wszystkie Tamary świata, ale wykrywanie prawidłowości w danych będzie możliwe dopiero wówczas, kiedy te dane zostaną zebrane, scalone i udostępnione w systemie.

Ogromna przepaść dzieli bogate, szczegółowe bazy znajdujące się w posiadaniu brokerów danych od szczupłych i rozproszonych baz, jakimi dysponują nasze służby zdrowia. Tymczasem dane medyczne to jeden wielki chaos. Nawet jeśli nasze szczegółowe dane medyczne są przechowywane w jednym miejscu (a często nie są), informacje te mają tak różną postać, że scalenie ich na potrzeby algorytmu jest dosłownie niemożliwe – są tam między innymi skany i zdjęcia, raporty i wykresy, zalecenia i zapiski. Odrębny problem stanowią informacje zapisane odręcznie – maszyna będzie musiała się nauczyć odczytywać pismo,

rozpoznawać wszystkie używane w nim skróty i akronimy, biorąc od razu poprawkę na ewentualne błędy. A nie doszliśmy jeszcze nawet do objawów! Czy jeśli pacjentka przeszła „grypę", to miała grypę prawdziwą czy tylko infekcję grypopodobną? Czy pacjent rzeczywiście „umiera z bólu", czy tylko trochę go kłuje w brzuchu? Medycyna jest nauką niewiarygodnie skomplikowaną i każdy kolejny poziom złożoności coraz bardziej utrudnia maszynie przetworzenie danych[46].

IBM to nie jedyny gigant giełdowy, który próbował coś zrobić z chaotyczną, nieustrukturyzowaną informacją medyczną. W 2016 roku firma DeepMind, wydział do spraw SI spółki Google, podpisała kontrakt z londyńską fundacją Royal Free NHS Trust. DeepMind uzyskał dostęp do danych medycznych z trzech miejskich szpitali w zamian za aplikację pomagającą lekarzom rozpoznawać ostre uszkodzenie nerek. W pierwotnym zamiarze uczące się algorytmy miały służyć wszechstronną pomocą służbie zdrowia, badacze zostali jednak zmuszeni do obniżenia poprzeczki i zadowolenia się znacznie prostszym celem; otrzymane dane okazały się niewystarczające i nie było po prostu szans na osiągnięcie założonego rezultatu.

Były to czysto praktyczne trudności w realizacji projektu, ale współpraca DeepMind z londyńską służbą zdrowia miała też inny, bardziej kontrowersyjny wymiar. Jako że badacze zobowiązali się alarmować lekarzy o uszkodzeniach nerek u ich pacjentów, a fundacja Royal Free nie miała żadnych danych dotyczących tych narządów, DeepMind uzyskał dostęp do c a ł e j dokumentacji, czyli kartotek około 1,6 miliona pacjentów, za ostatnie pięć lat.

Teoretycznie tak niewiarygodny zasób informacji mógłby uratować ogromną liczbę pacjentów. Ostre uszkodzenie nerek zabija tysiąc osób w ciągu miesiąca, i posiadając dane sięgające tak daleko wstecz, badacze DeepMind mieli szansę zidentyfikować ważne trendy historyczne. A ponieważ uszkodzenia nerek występują częściej u osób cierpiących na inne schorzenia, szeroka baza danych teoretycznie powinna ułatwić wykrycie sygnałów ostrzegawczych i związków z przyszłym stanem zdrowia pacjentów.

Nie było jednak szału radości – wieści o nowym projekcie wywołały falę oburzenia. Przyznanie firmie DeepMind pełnego dostępu do całej

kartoteki pacjenta oznaczało ujawnienie naprawdę w s z y s t k i c h informacji: kto i kiedy został przyjęty do szpitala; kto odwiedzał pacjenta podczas hospitalizacji; wyniki badań histopatologicznych i radiogramy; kto przerwał ciążę, kto miał depresję, a nawet u kogo zdiagnozowano AIDS. A najgorsze ze wszystkiego? Nikt nie spytał pacjentów o zgodę na udostępnienie ich danych, nie dał im możliwości wyboru, ba – nie powiadomił ich nawet, że biorą udział w projekcie[47].

Warto dodać, że firmie Google zabroniono wykorzystywać pozyskane informacje w działalności handlowej poza projektem. Trzeba przyznać, że Google ma dużo lepsze notowania w zakresie ochrony danych osobowych niż brytyjska służba zdrowia, której szpitale zostały unieruchomione w 2017 roku przez północnokoreański ransomware, ponieważ ich komputery ciągle pracowały na systemie Windows XP[48] – niemniej to bardzo niepokojące, że niezwykle potężna, wiodąca firma technologiczna ma dostęp do tego rodzaju poufnych informacji o pojedynczych obywatelach.

Problemy z prywatnością

Bądźmy szczerzy – firmie Google raczej nie brakuje wrażliwych, intymnych nawet informacji o każdym z nas. Instynkt jednak podpowiada, że nasze dane medyczne to zupełnie inna kategoria poufności. Osoby z czystą kartą zdrowia mogą tego nie rozumieć; gdybyśmy mieli wybierać między ujawnieniem światu swojej kartoteki medycznej a ujawnieniem historii przeglądania stron w internecie, na co byśmy się zdecydowali? Ja wiem, że bez zastanowienia wybrałabym to pierwsze, i przypuszczam, że wiele osób postąpiłoby tak samo. I nie chodzi o to, że mam coś szczególnie interesującego do ukrycia. Ale to pierwsze to zwykła historia mojej biologii, podczas gdy drugie to otwarta księga mojego charakteru.

Nawet jeśli dane medyczne są potencjalnie sprawą mniej wstydliwą, Timandra Harkness, autorka książki *Big Data: Does Size Matter?* i prezenterka programu *Future Proofing* emitowanego przez kanał 4 radia BBC, broni tezy, że stanowią one przypadek szczególny.

„Po pierwsze, dane medyczne wielu osób to jakby ich życiorys – powiedziała mi. – Na przykład co trzecia Brytyjka przeszła zabieg

usunięcia ciąży i pewnie by nie chciała, żeby wszyscy jej krewni i znajomi o tym wiedzieli". Timandra podkreśla również, że twoja kartoteka ma znaczenie nie tylko dla ciebie. „Po drugie, jeżeli ktoś zdobędzie twoje dane genetyczne, to dowie się czegoś także o twoich rodzicach, twoim rodzeństwie, twoich dzieciach". A od raz udostępnionych danych genetycznych nie ma już ucieczki. „Nie możesz zmienić swojej biologii ani się jej wyprzeć. Jeśli ktoś pobierze ci DNA, już nic nie zrobisz. Możesz zmienić sobie rysy twarzy, możesz nosić rękawiczki, żeby nie zostawiać odcisków palców, ale twoje DNA jest niezmienne. I na zawsze powiązane z tobą".

Dlaczego to takie ważne? Timandra w 2013 roku moderowała grupę fokusową; zwykłych ludzi pytano o to, co ich najbardziej martwi w fakcie gromadzenia ich danych medycznych. „Nie przejmowali się za bardzo tym, że ich dane mogą zostać wykradzione czy zhakowane. Bardziej niepokoiło ich to, że ktoś mógłby wyciągnąć jakieś wnioski o nich jako grupie, a potem wykorzystać je przeciwko nim.

Załóżmy, że udałoby się powiązać ich kartę lojalnościową z supermarketu z kartoteką medyczną. I kiedy teraz ktoś chciałby się poddać operacji biodra, lekarz mógłby powiedzieć: «O, przykro mi, ale widzę tutaj, że kupowała pani ostatnio sporo pizzy i kupowała pani mnóstwo papierosów, więc obawiam się, że musimy przesunąć panią na koniec kolejki»".

Nie są to obawy bezpodstawne, zwłaszcza w Wielkiej Brytanii, gdzie niektóre zubożałe szpitale państwowej służby zdrowia w kolejkach do operacji kolana i biodra już dają preferencje pacjentom niepalącym[49]. Jest też na świecie wiele państw, gdzie pacjentom z otyłością odmawia się ubezpieczenia lub leczenia[50].

Sprawa stanowi pewien paradoks. Moglibyśmy ogromnie skorzystać na udostępnieniu naszych kartotek medycznych algorytmom. Watson firmy IBM wcale nie musiałby pozostawać robotem z filmów fantastycznych. Żeby jednak ziścić nasze marzenia o lepszym zdrowiu, musielibyśmy oddać nasze dane najbogatszym koncernom świata, bo tylko one będą w stanie przeprowadzić nas przez pole minowe wyzwań stojących przed konstruktorami naszego upragnionego robota medycznego. A zrzekając się prywatności, zawsze musimy liczyć się z tym, że nasze dane mogą

zostać sfałszowane, skradzione lub użyte przeciwko nam. Czy jesteśmy gotowi na takie ryzyko? Czy wierzymy w algorytmy i korzyści płynące z ich zastosowania na tyle mocno, by poświęcić swoją prywatność?

Genetyczna wpadka

Francis Galton był wiktoriańskim statystykiem i genetykiem, jednym z najwybitniejszych przedstawicieli swojego pokolenia – i dalekim kuzynem Karola Darwina. Wiele jego idei wywarło głęboki wpływ na współczesną naukę – wystarczy wspomnieć, że prace Galtona stanowiły praktycznie fundament nowoczesnej statystyki. Mamy za to wszystko wobec niego ogromny dług wdzięczności. (Niestety, działał także aktywnie w prężnie rozwijającym się wówczas ruchu eugenicznym, za co na pewno pomnika mu nie postawimy).

Galton chciał badać ludzkie cechy za pośrednictwem danych i już wówczas wiedział, że jeżeli chce się odkryć coś godnego uwagi, potrzeba tych informacji naprawdę dużo. Wiedział również, że ludzie przejawiają ogromną ciekawość wobec własnego ciała. Odkrył, że ów nienasycony apetyt, jeśli właściwie go pobudzić, potrafi wziąć górę nad potrzebą prywatności. Mało tego – ludzie częstokroć byli wręcz gotowi zapłacić za możliwość zaspokojenia swojego wilczego głodu wiedzy.

Galton zwietrzył okazję w 1884 roku, kiedy pod patronatem królowej Wiktorii otwarto w Londynie ogromną wystawę dokumentującą postępy opieki zdrowotnej w Wielkiej Brytanii. Za własne pieniądze wynajął stoisko, które nazwał Laboratorium Antropometrycznym – miał nadzieję, że wśród milionowej publiczności znajdzie się kilka osób, które zechcą zapłacić za pomiary własnego ciała.

Znalazło się ich więcej niż kilka. Klienci walili do niego drzwiami i oknami – ludzie aż się palili, żeby, lżejsi o trzy pensy, wkroczyć do laboratorium Galtona. W środku każdy mógł się zmierzyć ze specjalnie skonstruowanymi przyrządami, poddając sprawdzianowi między innymi ostrość wzroku, „rzut oka" (rozpoznawanie kształtów, rozpoznawanie środka linii), siłę uciągu i ścisku oraz szybkość uderzenia[51]. Laboratorium było tak oblegane, że Galton z konieczności zaczął wpuszczać po dwie osoby naraz. (Szybko zorientował się, że dzieci i rodziców najlepiej

sprawdzać osobno, żeby nie tracić czasu na leczenie zranionej miłości własnej tych drugich. W artykule napisanym po wystawie skomentował to tak: „Starszym nie podobało się, że młodsi na próbach wypadają lepiej od nich, nalegali więc na ich powtórzenie”[52]).

Bez względu na to, czy wyniki pomiarów były bardzo dobre, czy bardzo złe, Galton zapisywał je na białej karcie, którą wręczał badanym na pamiątkę. W rzeczywistości jednak największym zwycięzcą był sam uczony: kończył wystawę z kopiami wszystkich pomiarów – bezcennym zestawem danych biometrycznych 9337 osób – i ciężkim mieszkiem.

Przeskoczmy szybciutko 130 lat do przodu, a być może dostrzeżemy pewne podobieństwa we współczesnej modzie na testy genetyczne. Za śmieszną cenę 149 funtów możesz wysłać próbkę śliny firmie 23andMe, która zajmuje się genomiką i biotechnologią, i otrzymać swoją charakterystykę genetyczną oraz odpowiedzi na takie pytania: „Jaki masz typ woskowiny?”, „Czy masz geny zrośniętych brwi?”[53]. A może gen powodujący kichanie, kiedy patrzysz na słońce?[54] Są też pytania poważniejsze: „Czy masz skłonności do raka piersi?”, „Czy masz genetyczne predyspozycje do choroby Alzheimera?”[55].

W ten sprytny sposób firma 23andMe zgromadziła potężną bazę danych genetycznych, liczącą dziś miliony próbek. Praktyka nie różniła się niczym od tego, co giganci internetu robią od dawna, należy jednak pamiętać, że oddając w ramach wymiany nasze DNA, udostępniamy komuś najbardziej osobiste dane, jakimi dysponujemy. Z tych danych powstała baza, z której wszyscy będziemy czerpać korzyści – niezwykle cenny zasób, mogący pogłębić nasze rozumienie ludzkiego genomu. Naukowcy, firmy farmaceutyczne i organizacje non profit z całego świata już ustawiają się w kolejce do firmy 23andMe, żeby szukać prawidłowości w zgromadzonych przez nią danych w nadziei znalezienia odpowiedzi na wielkie pytania, które odmienią życie nas wszystkich: „Jakie są dziedziczne przyczyny chorób?”, „Czy można wynaleźć nowe leki na określone choroby?”, „Czy istnieje lepszy sposób na leczenie choroby Parkinsona?”.

Ta baza danych jest bardzo cenna także w dosłownym znaczeniu tego słowa. Mimo że projekt 23andMe ma ogromne znaczenie dla społeczeństwa, firma nie prowadzi swoich badań z dobrego serca. Za

zgodą swoich klientów (80 procent przypadków) odsprzedaje anonimowe wersje ich danych genetycznych swoim partnerom badawczym[56]. I to za niemałe pieniądze, które bynajmniej nie stanowią dodatkowego bonusu dla tej firmy – w rzeczywistości to podpora ich planu biznesowego. Jeden z członków zarządu spółki 23andMe powiedział w wywiadzie dla „Fast Company": „Naszym długofalowym celem nie jest sprzedaż zestawów do pobierania próbek, choć są one niezbędne do pozyskania podstawowych danych". Warto o tym pamiętać, kiedy zamówisz sobie płatny raport genetyczny: to nie ty używasz produktu – to ty nim jesteś[57].

Słowo ostrzeżenia. Nie ufałabym też w pełni obietnicom zachowania anonimowości. W 2005 roku pewien młody mężczyzna[58], który został poczęty z nasienia anonimowego dawcy, zdołał zidentyfikować swojego biologicznego ojca – wysłał do analizy próbki swojej śliny i wykorzystał wskazówki zawarte w swoim kodzie DNA[59]. A w 2013 roku grupa naukowców opublikowała słynny już dziś artykuł, w którym zademonstrowała, że za pomocą genów można zidentyfikować miliony osób, i to używając do tego wyłącznie komputera osobistego oraz kilku dobrze sformułowanych wyszukiwań w internecie[60].

Jest jeszcze jeden powód, żeby nie umieszczać swojego DNA w żadnej bazie danych. Choć istnieje prawo chroniące obywateli przed najgorszymi, bo genetycznymi, prześladowaniami – nie zmierzamy więc raczej ku przyszłości, w której Beethoven czy Stephen Hawking będą oceniani na podstawie swoich predyspozycji genetycznych, a nie talentu – zapisy te nie dotyczą ubezpieczycieli. Nikt nie może pobrać od ciebie próbki do badania DNA, jeśli nie wyrazisz na to zgody, ale w Stanach Zjednoczonych ubezpieczyciel może spytać klienta, czy poddał się badaniu oceniającemu ryzyko zachorowania na takie choroby, jak parkinson, alzheimer czy rak piersi, i odmówić podpisania umowy ubezpieczenia na życie, jeśli klient udzieli odpowiedzi, która mu się nie spodoba. W Wielkiej Brytanii firmy ubezpieczeniowe mają prawo uwzględnić w swoich kalkulacjach wyniki badania DNA pod kątem choroby Huntingtona (jeśli suma ubezpieczenia przekracza 500 tysięcy funtów)[61]. Oczywiście można próbować kłamać i udawać, że nigdy takiego testu się nie robiło, ale ujawnione poświadczenie nieprawdy automatycznie unieważnia umowę. Tak więc

jedynym sposobem uniknięcia podobnej dyskryminacji jest niepoddawanie się nigdy temu badaniu. Czasem niewiedza naprawdę jest błogosławieństwem.

Dla zrozumienia zdrowia człowieka nie można sobie wyobrazić niczego cenniejszego od zsekwencjonowania genomów milionów ludzi. Mimo to w najbliższej przyszłości raczej nie zamierzam się poddawać badaniom genetycznym, choć (na szczęście dla społeczeństwa) miliony osób dobrowolnie przesyłają swoje próbki do takich badań. Według ostatnich danych firma 23andMe ma ponad 2 miliony klientów z opisanym genomem[62], a MyHeritage, Ancestry.com czy projekt National Geographic zgromadziły ich kolejne miliony. Być może więc mamy do czynienia z paradoksem, który wcale paradoksem nie jest. Ostatecznie rynek przemówił: oddanie swojej prywatności za szansę dorzucenia cegiełki do budowy lepszego społeczeństwa nie brzmi szczególnie atrakcyjnie, ale już za informację, że w 25 procentach jesteś wikingiem – proszę bardzo![13*]

Najwyższe dobro?

Oczywiście żartuję. Nie można od nikogo oczekiwać, że podejmując decyzję, czy wysłać swoją próbkę do badań genetycznych, będzie myśleć przede wszystkim o wielkich wyzwaniach czekających służbę zdrowia w bliższej i dalszej przyszłości – to byłoby nierozsądne. Wręcz przeciwnie – nietrudno zrozumieć, dlaczego ludzie tego nie robią; jako jednostki mamy inne priorytety względem siebie, a inne względem całej ludzkości.

To jednak wskazuje nam ostatnie ważne zagadnienie. Gdyby rzeczywiście dało się zbudować maszynę diagnostyczną zdolną do zalecania najlepszego leczenia, komu miałaby ona służyć? Jednostce czy populacji? Bo będą takie przypadki, kiedy trzeba będzie między nimi wybierać.

Wyobraźmy sobie na przykład, że idziemy do przychodni z wyjątkowo nieprzyjemnym kaszlem. Nic poważnego, pewnie samo by przeszło, ale trafiliśmy na maszynę, która tak na wszelki wypadek wolałaby posłać nas na prześwietlenie i badanie krwi. Pewnie przepisałaby też antybiotyki, gdybyśmy o nie poprosili. Choćby miały skrócić nasze męki tylko o parę

dni, to gdyby jedynym celem maszyny były nasze zdrowie i wygoda, algorytm mógłby uznać, że gra jest warta świeczki.

Gdyby jednak maszyna została skonstruowana z myślą o całej populacji, byłaby znacznie ostrożniejsza w obliczu zagadnień związanych z odpornością na antybiotyki. Jeśli tylko nasze życie nie byłoby bezpośrednio zagrożone, to nasz dyskomfort nie miałby dla algorytmu żadnego znaczenia – wycisnęlibyśmy z niego receptę na antybiotyki tylko w obliczu najwyższej konieczności. Maszyna mogłaby też nie być szczególnie skłonna marnować na nas środki finansowe, wiedziałaby na pewno o długich kolejkach do specjalistów i badań, więc nie posłałaby nas na dalszą diagnostykę, jeśli nie mielibyśmy objawów wskazujących na coś poważniejszego. Szczerze mówiąc, pewnie kazałaby nam zażyć aspirynę i przestać się nad sobą rozczulać.

Podobnie algorytm mający na uwadze dobro społeczne mógłby przy podejmowaniu decyzji, komu przydzielić pobrany od dawcy organ, kierować się zasadą „ratowania możliwie największej liczby ludzi". A to mogłoby zaowocować zupełnie innym planem leczenia od tego, jaki stworzyłaby maszyna kierująca się w działaniu wyłącznie twoim dobrem.

Algorytm pracujący dla państwowej służby zdrowia czy firmy ubezpieczeniowej starałby się pewnie maksymalnie ciąć koszty, podczas gdy maszyna zaprojektowana dla koncernu farmaceutycznego sugerowałaby użycie wyprodukowanego przez nią leku.

Zastosowania na polu medycyny są zapewne mniej kontrowersyjne niż przykłady z wymiaru sprawiedliwości – tutaj nie ma obrony i nie ma oskarżyciela. Wszystkim pracownikom służby zdrowia przyświeca jeden i ten sam cel: poprawić stan zdrowia pacjenta. Lecz nawet na tym polu występują subtelne różnice w dążeniach zainteresowanych stron.

W każdej dziedzinie życia, do której wprowadzimy algorytmy, ustali się z czasem stan równowagi – między prywatnością a wyższym dobrem, jednostką a społeczeństwem, wyzwaniami a priorytetami. Niełatwo odnaleźć drogę w plątaninie różnych motywów, nawet gdy na jej końcu widać wyraźnie wizję lepszej opieki zdrowotnej dla nas wszystkich.

Tym trudniej to zrobić, kiedy sprzeczne ze sobą motywacje pozostają ukryte. Kiedy pożytki z algorytmu są gloryfikowane, a zagrożenia

ukrywane w głębokim cieniu. Kiedy zastanawiamy się, komu i w co wierzyć. Oraz kto na tym skorzysta, jeżeli uwierzymy.

Samochody

Był świt 13 marca 2004 roku i słońce ledwo zdążyło się wznieść nad horyzont, a bar Slash X w głębi pustyni Mojave już pękał w szwach[1]. Bar ów znajduje się na obrzeżach niewielkiego miasta Barstow, położonego mniej więcej w połowie drogi między Los Angeles i Las Vegas – niedaleko miejsca, w którym Uma Thurman wykopywała się z trumny na planie zdjęciowym *Kill Billa II*[2]. Slash X odwiedzają z reguły kowboje i miłośnicy bezdroży, ale w tamten wiosenny dzień oblegała go zupełnie inna klientela. Wzniesiona przed budynkiem prowizoryczna widownia nie mogła pomieścić podekscytowanej publiczności oraz szalonych konstruktorów i fanatyków motoryzacji, których przygnało tam jedno marzenie: znaleźć się w gronie pierwszych ludzi na Ziemi, którzy widzieli, jak samochód bez kierowcy wygrywa wyścig na torze.

Imprezę zorganizowała DARPA – Agencja Zaawansowanych Projektów Badawczych Departamentu Obrony Stanów Zjednoczonych (ang. *US Defence Advanced Research Projects Agency*), lepiej znana w Pentagonie jako „wydział szurniętych naukowców"[3]. DARPA interesowała się bezzałogowymi pojazdami już od pewnego czasu i nie bez powodu: miny przeciwtransportowe i ataki na pojazdy wojskowe powodują znaczne straty w ludziach. Na początku 2004 roku przedstawiciele Agencji oświadczyli, że do roku 2015 DARPA zamierza zautonomizować jedną trzecią naziemnych pojazdów wojskowych[4].

Ten ambitny projekt miał już swoje lata i choć posuwał się do przodu żółwim tempem, kosztował niemało: w ciągu dwudziestu lat DARPA wydała mniej więcej pół miliarda dolarów na finansowanie badań akademickich i prac badawczych w prywatnych firmach[5]. W końcu ktoś wpadł na genialny pomysł: zorganizujmy zawody! No i Agencja zaprosiła wszystkich zainteresowanych z całego kraju – konstruktorów samochodów bez kierowcy – do otwartego wyścigu na długim torze, z nagrodą miliona

dolarów dla zwycięzcy[6]. Miał to być pierwszy tego typu wyścig na świecie, a przy tym tani sposób na szybki postęp w realizacji planów Agencji.

Tor liczył 227 kilometrów i DARPA zadbała o to, żeby nie był łatwy – trasa przebiegała przez strome podjazdy i zjazdy, skały, parowy i nierówności, a nawet kolonie kaktusów; nie zabrakło też wąskich polnych dróżek mających niekiedy zaledwie metr szerokości. Dwie godziny przed startem organizatorzy wręczyli każdemu zespołowi płytę CD z koordynatami GPS[7]. Był to zbiór dwóch tysięcy punktów orientacyjnych rozrzuconych po całej trasie jak okruszki chleba – akurat tyle, żeby dać autonomicznym uczestnikom wyścigu ogólne pojęcie o przebiegu toru, za mało jednak na dokładną nawigację między czyhającymi na trasie przeszkodami.

Trudności były zniechęcające, niemniej zgłosiło się 106 nieustraszonych zespołów. Kwalifikacje przeszło 15 pojazdów, które uznano za wystarczająco bezpieczne. Były wśród nich pojazdy wyglądające jak łaziki pustynne, były potwory przypominające monster trucki i samochody, które można było łatwo pomylić z czołgiem. Jeden z uczestników wziął podobno kredyt hipoteczny, żeby skonstruować swój pojazd. Inny umieścił na dachu swego samochodu dwie deski surfingowe – żeby się wyróżniać. Startował nawet jeden motocykl z automatyczną pionizacją[8].

Rankiem w dniu zawodów na oczach kilkutysięcznej widowni przy barze Slash X zebrało się kilkanaście autonomicznych samochodów i jeden motocykl bez kierowcy. Pojazdy podjeżdżały po kolei do linii startu, a każdy kolejny coraz bardziej przenosił widzów w świat *Mad Maksa* albo *Wacky Races*. Ale nie wygląd miał tu znaczenie. Liczyło się tylko jedno: żeby przejechać cały tor w czasie nie dłuższym niż dziesięć godzin.

Na niespodzianki nie trzeba było długo czekać. Jeden samochód dachował już w strefie startu i został wycofany z wyścigu[9]. Motocykl przewrócił się tuż za linią startu i został zdyskwalifikowany. Inny pojazd uderzył w betonowy mur po przejechaniu niecałych 50 metrów. Jeszcze inny ugrzązł w zasiekach z drutu kolczastego. Kolejny zablokował się między dwiema roślinami, biegaczami pustynnymi, bo uznał je za przeszkody nieruchome, więc bezustannie cofał i ruszał do przodu i nie uwolniłby się z tej pułapki bez interwencji człowieka[10]. Pozostałe pojazdy ciągle zderzały się ze skałami i wjeżdżały do rowów. Pękały osie, rozrywały

się opony, w powietrze wzbijały się płaty karoserii[11]. Teren wokół baru Slash X zaczął przypominać cmentarzysko robotów.

Najlepszy wynik uzyskał pojazd wystawiony przez zespół z Carnegie Mellon University – zdołał przejechać imponujące 11,9 kilometra, zanim źle ocenił teren i zawiesił się na podwoziu, a ponieważ w pojeździe nie było człowieka, który mógłby zaradzić sytuacji, koła buksowały tak długo, aż w końcu pojazd stanął w płomieniach[12]. O jedenastej przed południem było już po wszystkim. Przedstawiciel Agencji poleciał helikopterem na linię mety, żeby poinformować czekających dziennikarzy, iż żaden z pojazdów tam nie dotrze[13].

Rozlany olej, kurz, hałas i ogólna destrukcja – tak wyglądał pierwszy, nierozstrzygnięty wyścig zorganizowany przez DARPA. Wszystkie zespoły przez rok trudziły się nad stworzeniem pojazdu, który w najlepszym razie wytrzymał na torze przez kilka minut. Nie oznacza to jednak, że wyścig był katastrofą. Rywalizacja okazała się prawdziwym inkubatorem nowych idei i podczas kolejnego Grand Challenge rok później było widać gołym okiem prawdziwy skok technologiczny: wszyscy poza jednym uczestnikiem pobili 12-kilometrowy próg z 2004 roku. A najbardziej zdumiewające było to, że aż pięć różnych samochodów zdołało pokonać cały dystans 212,4 kilometra bez jakiejkolwiek interwencji człowieka[14].

Dziś, po zaledwie dekadzie z okładem, zewsząd dochodzą głosy, że przyszłość transportu to pojazdy autonomiczne. Pod koniec 2017 roku Philip Hammond, brytyjski kanclerz skarbu, oświadczył, że rząd planuje wprowadzić na drogi w pełni autonomiczne samochody – poruszające się bez kontroli człowieka – już w roku 2021. Daimler obiecał, że będzie miał samochody autonomiczne w 2020 roku[15], a Ford w 2021[16]; pozostali producenci również przedstawili własne podobne prognozy.

Polemiki w prasie przestały już dotykać tematu realności tych wizji; teraz debatuje się o problemach, z którymi będziemy musieli się zmierzyć, kiedy autonomiczne samochody w końcu się pojawią. „Czy twój autonomiczny samochód ma przejechać przechodnia, żeby cię ratować?", zapytywał „New York Times" w czerwcu 2016 roku[17], a w listopadzie 2017: „Co się stanie z przejechanymi zwierzętami i z mandatami, kiedy nasze pojazdy zaczną same jeździć?"[18]. Tymczasem w styczniu 2018 roku

„Financial Times" ostrzegał: „Transport ciężarowy zmierza ku autonomicznej przyszłości: związki alarmują, że miliony kierowców pozostaną bez pracy"[19].

Cóż więc się zmieniło? Jakim cudem ta technologia zdołała przejść transformację z nieporadnej niekompetencji do rewolucyjnej pewności siebie w ciągu kilku zaledwie lat? I czy naprawdę możemy liczyć na to, że ów błyskawiczny postęp to długofalowy trend?

Co ja widzę?

Nasze marzenie o doskonałym pojeździe autonomicznym poczęło się w dawnych czasach, w erze fantastycznonaukowych rojeń o plecakach odrzutowych, rakietach, skafandrach kosmicznych z folii aluminiowej i broni laserowej. Na nowojorskiej Wystawie Światowej w 1939 roku firma General Motors objawiła własną wizję przyszłości. Goście mogli usiąść w zamontowanych na taśmie transportowej „fotelach dźwiękowych" i wybrać się w szesnastominutową podróż do świata fantazji[20]. Za szybą mogli obejrzeć makietę wielkiego marzenia amerykańskich konstruktorów: przecinające kraj wzdłuż i wszerz superautostrady, drogi łączące farmy z miastami, ulice i skrzyżowania pełne automatycznych, sterowanych falami radiowymi samochodów zdolnych przemieszczać się bezpiecznie z prędkością dochodzącą do 160 kilometrów na godzinę. „Dziwne? – pytał fotel dźwiękowy. – Fantastyczne? Niewiarygodne? Pamiętajmy, że to świat z początku lat sześćdziesiątych!"[21].

Wielokrotnie podejmowano próby urzeczywistnienia tego marzenia. Sam koncern General Motors przymierzał się do tego już w latach pięćdziesiątych, wypuszczając na rynek model Firebird II[22]. W latach sześćdziesiątych brytyjscy naukowcy próbowali przerobić citroena DS19 tak, żeby mógł komunikować się z drogą (pod autostradą M4, między miejscowościami Slough i Reading, do dziś może leżeć dziewięciomilowy odcinek kabla elektrycznego, pozostałość po tamtych eksperymentach)[23]. W latach osiemdziesiątych był projekt Navlab, nad którym pracowali badacze z Carnegie Mellon University. W latach dziewięćdziesiątych Unia Europejska wydała miliard euro na Eureka Prometheus Project[24]. Przy

każdym nowym projekcie spełnienie marzenia o samochodzie bez kierowcy zdawało się już tak blisko, niemalże na wyciągnięcie ręki.

W pierwszej chwili można by sądzić, że zbudowanie samochodu bez kierowcy powinno być stosunkowo łatwe; ostatecznie większość ludzi potrafi opanować umiejętności niezbędne do prowadzenia pojazdu. Poza tym są tylko dwie możliwe zmienne: prędkość i kierunek – trzeba tylko wiedzieć, jak mocno wcisnąć pedał i o ile skręcić kierownicę. Co w tym trudnego?

A jednak pierwszy wyścig z cyklu DARPA Grand Challenge wykazał niezbicie, że zbudowanie pojazdu autonomicznego jest w rzeczywistości znacznie trudniejsze, niż mogłoby się wydawać. Okazuje się, że kiedy próbujesz zaprogramować algorytm, który ma kontrolować kupę żelastwa pędzącą z prędkością 100 kilometrów na godzinę, sprawy szybko zaczynają się komplikować.

Na przykład sieci neuronowe, które z dużym powodzeniem wykorzystuje się do wykrywania guzów w tkance piersiowej, powinny się idealnie nadawać do wspierania systemu sterowania pojazdu autonomicznego w „widzeniu" otoczenia. Sieci neuronowe (choć w formie nieco okrojonej w porównaniu z najnowocześniejszymi wersjami) montowano w prototypowych pojazdach bez kierowcy już w 2004 roku[25], starając się je nauczyć rozumienia obrazów przesyłanych z kamer dachowych. Można z nich wyłuskać bardzo wiele cennych informacji – sieć neuronowa rozumie kolory, teksturę, a nawet cechy fizyczne filmowanego wycinka rzeczywistości, takie jak linie proste i zakrzywione, krawędzie i kąty. Tylko co potem zrobić z tymi wszystkimi informacjami?

Załóżmy, że rozkażesz swojemu samochodowi: „Jedź tylko po tym, co wygląda jak asfalt" – tak sformułowana komenda nie sprawdziłaby się jednak na pustyni, gdzie wszystkie drogi są gruntowe. Powiesz mu zatem: „Jedź po najgładszej powierzchni" – ale i tutaj czeka porażka: najgładszą powierzchnią jest niemal zawsze niebo albo szklana ściana budynku. Można by więc odwołać się do zupełnej abstrakcji przy opisywaniu drogi: „Szukaj przedmiotu o dwóch mniej więcej prostych krawędziach. Linie powinny być szeroko rozstawione na dole obrazu i zbliżać się do siebie na górze obrazu". Taki opis wydaje się całkiem sensowny, jest jednak pewien

szkopuł: dokładnie tak samo wygląda drzewo na zdjęciu, a namawianie samochodu do wjeżdżania w drzewa to raczej nie najlepszy pomysł.

Kłopot z kamerami jest taki, że nie dają poczucia skali ani odległości. Cechę tę z powodzeniem wykorzystują reżyserowie w swoich filmach – pomyśl chociaż o scenie otwierającej *Gwiezdne wojny*, w której niszczyciel gwiezdny powoli wyłania się na tle atramentowoczarnej przestrzeni kosmicznej i zawisa złowrogo w górnej części kadru. Masz wrażenie, że to gigant, kiedy w rzeczywistości filmowano model o długości niecałego metra. Trik daje świetny efekt na dużym ekranie. Ale w samochodzie bez kierowcy, kiedy dwie cienkie linie równoległe mogą być zarówno ciągnącą się po horyzont drogą przed pojazdem, jak i pniem stojącego przy jezdni drzewa, dokładna ocena odległości staje się sprawą życia i śmierci.

Nawet jeśli użyjesz większej liczby kamer i sprytnie połączysz pochodzące z nich dane w trójwymiarowy obraz otoczenia pojazdu, zostanie do rozwiązania jeszcze jeden potencjalny problem. Problem ten odkrył w latach dziewięćdziesiątych Dean Pomerleau, badacz z Carnegie Mellon University. Pracował nad pojazdem o nazwie ALVINN – był to skrótowiec utworzony od angielskiej nazwy *Autonomous Land Vehicle In a Neural Network* (Autonomiczny pojazd drogowy w sieci neuronowej). ALVINN-a ćwiczono w rozumieniu otoczenia na podstawie zachowań kierującego pojazdem człowieka. Pomerleau i jego współpracownicy siadali za kierownicą i zabierali ALVINN-a w długie trasy, rejestrując wszystkie swoje czynności. Zgromadzone w ten sposób dane wykorzystywano do uczenia sieci neuronowej: jedź wszędzie tam, gdzie jeździ człowiek, całej reszty unikaj[26].

Z początku procedura sprawdzała się świetnie. Po treningu ALVINN potrafił jechać samodzielnie po nieskomplikowanej drodze. Ale kiedy dojechał do mostu, zrobiło się groźnie – samochód nagle niebezpiecznie skręcił i Pomerleau musiał złapać kierownicę, żeby nie doszło do kraksy.

Po kilku tygodniach analizowania danych z miejsca incydentu Pomerleau znalazł problem: wszystkie drogi, na których ćwiczył ALVINN, miały trawiaste pobocza. Podobnie jak w przypadku sieci neuronowych w rozdziale *Medycyna*, które rozpoznawały wilki na podstawie śniegu w tle zdjęcia, dla sieci neuronowej ALVINN-a kluczowym wyznacznikiem toru

jazdy była trawa po obu stronach drogi. Kiedy tylko trawa zniknęła, komputer zupełnie się pogubił.

Lasery w odróżnieniu od kamer potrafią mierzyć odległość. Pojazdy wyposażone w system o nazwie LiDAR (*Light Detection and Ranging*, Laserowe wykrywanie i pomiar odległości), po raz pierwszy zastosowane w drugim wyścigu DARPA Grand Challenge w 2005 roku, emitują z lasera fotony i na podstawie czasu ich powrotu po odbiciu się od przeszkody obliczają dystans dzielący od niej pojazd. Ale jest i druga strona medalu: LiDAR nie potrafi rozróżniać kolorów ani tekstur, jest beznadziejny w czytaniu znaków drogowych i niespecjalnie dobry w pomiarze większych odległości. Z kolei radar – ta sama zasada, tylko przy wykorzystaniu fal radiowych – dobrze sobie radzi przy każdej pogodzie, potrafi wykrywać odległe przeszkody, a nawet widzi na wskroś niektórych substancji, lecz jest kompletnie bezużyteczny, jeśli chodzi o bardziej szczegółowe informacje o kształcie czy strukturze przeszkody.

Samodzielnie żadne z tych źródeł danych – kamera, LiDAR, radar – nie pozwoli algorytmowi zrozumieć, co się dzieje dookoła pojazdu. A zatem cała sztuka polega na umiejętnym ich połączeniu. Byłoby to stosunkowo proste, gdyby urządzenia te zgadzały się co do tego, co widzą, ale ponieważ tak nie jest – sprawa robi się o wiele bardziej złożona.

Weźmy przykład z biegaczami pustynnymi, które zablokowały jeden z samochodów na pierwszym wyścigu DARPA Grand Challenge. Wyobraźmy sobie, że to nasz autonomiczny pojazd znalazł się w tak niefortunnym położeniu. LiDAR mówi mu, że ma przed sobą przeszkodę. Kamera się zgadza. Radar, którego fale przenikają przez kulę suchych gałązek, informuje, że nie widzi żadnego problemu. Któremu czujnikowi powinien zaufać algorytm?

Co będzie, jeśli kamera otrzyma najwyższy priorytet? Wyobraź sobie, że w pochmurny dzień przecina ci drogę wielka biała ciężarówka. Tym razem LiDAR i radar są zgodne – trzeba natychmiast hamować, kamera natomiast nie dostrzega niczego groźnego, bo ciężarówka zlewa się z białą monotonią chmur.

Do tego wszystkiego dochodzi jeszcze jeden problem. Nie dość, że czujniki mogą fałszywie zinterpretować otoczenie, to trzeba wziąć poprawkę i na tę ewentualność, że mogą się po prostu pomylić.

Pewnie widzieliście nieraz na Google Maps wielki niebieski krąg wokół waszej lokalizacji – jego obecność sygnalizuje potencjalny błąd w pomiarze GPS. Kiedy twoja lokalizacja zostanie ustalona z dużą dokładnością, krąg będzie miał postać małej kropki, zdarza się jednak i tak, że obejmuje on znacznie większy obszar, a środek okręgu zostaje ustawiony w niewłaściwym miejscu. Najczęściej nie ma to większego znaczenia. Wiemy, gdzie się znajdujemy, i możemy odrzucić błędną informację. Ale samochód bez kierowcy nie ma takiej empirycznej płaszczyzny odniesienia i nie ma pojęcia o własnym położeniu. Poruszając się po pasie ruchu o szerokości niecałych czterech metrów, nie może polegać na samym GPS-ie, jeśli ma dokładnie wiedzieć, gdzie się znajduje.

GPS to nie jedyne urządzenie, które może podawać niezbyt ścisłe informacje. Każdy pomiar wykonany przez sprzęt zainstalowany w samochodzie autonomicznym będzie obarczony marginesem błędu – czy będzie to odczyt z radaru, pochylenie wzdłużne i poprzeczne, obroty kół czy bezwładność pojazdu. Nic nie jest nigdy pewne na sto procent. Na dodatek warunki zewnętrzne potrafią jeszcze pogorszyć sytuację: deszcz wpływa na dokładność LiDAR-u[27], ostre słońce na kamery[28], a długa jazda po wybojach potrafi zniszczyć akcelerometry[29].

Rezultatem jest jeden wielki chaos sygnałowy. Pytania z pozoru proste: „Gdzie jesteś?”, „Co się znajduje dookoła?”, „Co robić?” – nagle stają się przerażająco trudne. Którym informacjom wierzyć? Udzielenie odpowiedzi na to pytanie jest już prawie niemożliwe.

Właśnie – p r a w i e niemożliwe.

Na szczęście można z tego chaosu wybrnąć, ponieważ istnieje sposób na snucie sensownych domysłów w nieuporządkowanym świecie. A wszystko dzięki potężnej formule matematycznej o fenomenalnych możliwościach, która przeszła do historii jako twierdzenie Bayesa.

Wielki Kościół Czcigodnego Bayesa

Nie jest przesadą, że twierdzenie Bayesa to jedna z najbardziej wpływowych idei w historii człowieka. Wśród naukowców, ekspertów od uczenia maszynowego i statystyków twierdzenie to obrosło swoistym

kultem. A przecież sama idea jest niewiarygodnie prosta. Tak prosta, że z początku można by pomyśleć, iż pan Bayes stwierdził jedynie rzecz oczywistą.

Postaram się wyjaśnić rzecz na banalnym przykładzie.

Wyobraź sobie, że jesz kolację w restauracji. W pewnym momencie twoja przyjaciółka pochyla się do ciebie i szepcze, że przy stoliku za twoimi plecami siedzi Lady Gaga.

Zanim się obejrzysz, skalkulujesz sobie, w jakim stopniu możesz wierzyć przyjaciółce. Weźmiesz pod uwagę swoją dotychczasową wiedzę – takie czynniki, jak jakość lokalu, odległość, jaka dzieli cię od domu Gagi w Malibu, wzrok twojej towarzyszki. Gdyby ktoś bardzo nalegał, twoje przekonanie dałoby się wyrazić liczbowo, pewnym stopniem prawdopodobieństwa.

Kiedy już się obrócisz i spojrzysz na tę rzekomą Lady Gagę, wszystko, co zarejestrują twoje oczy, zostanie automatycznie wykorzystane do wzmocnienia lub osłabienia twojego przekonania. Być może platynowe blond włosy będą się zgadzać ze spodziewanym obrazem Lady Gagi, tak więc twoja pewność wzrośnie. Ale fakt, że kobieta siedzi sama, bez ochroniarzy, już tę pewność osłabi. Każde kolejne spostrzeżenie będzie budować końcową ocenę.

Tak właśnie działa twierdzenie Bayesa: daje nam metodę systematycznej aktualizacji naszego przekonania o słuszności określonej hipotezy na podstawie dostępnych informacji[30]. I choć Bayes potwierdza, że nigdy nie możemy mieć całkowitej pewności co do słuszności rozpatrywanej teorii, pozwala nam jednak wybrać – w świetle zaprezentowanych dowodów – najlepszy z możliwych domysłów. Kiedy więc zobaczysz, że kobieta przy sąsiednim stoliku ma na sobie suknię zrobioną z płatów mięsa, ta informacja może okazać się wystarczająca, żeby przekroczyć próg pewności i dojść do wniosku, że to rzeczywiście Lady Gaga.

Twierdzenie Bayesa nie jest jednak zwykłą formułą, która pokazuje, jak ludzie podejmują decyzje. Jego znaczenie jest o wiele większe. Przytoczę w tym miejscu słowa Sharon Bertsch McGrayne, autorki książki *The Theory That Would Not Die* („Teoria, która nie chce umrzeć”): „Bayes podważa głęboko zakorzenione przeświadczenie, że nowoczesna nauka wymaga

obiektywizmu i precyzji”[31]. Wyposażywszy nas w mechanizm umożliwiający pomiar naszej wiary, Bayes pozwala nam wyciągać rozsądne wnioski z pobieżnych obserwacji, z nieuporządkowanych, niepełnych i przybliżonych danych – a nawet z niewiedzy.

Ale jego twierdzenie nie służy nam wyłącznie do potwierdzenia przeczuć. Okazuje się, że kiedy jesteśmy zmuszeni do kwantyfikacji przekonań, często dochodzimy do wniosków sprzecznych z tym, co podpowiada nam intuicja. Twierdzenie Bayesa wyjaśnia, dlaczego więcej kobiet niż mężczyzn rozpoznaje się fałszywie jako przyszłych morderców – jak w przykładzie na stronie 92, w rozdziale *Wymiar sprawiedliwości*. I to twierdzenie Bayesa wyjaśnia, dlaczego jest tak, że choć ktoś otrzymuje diagnozę „rak piersi”, testy mogą być obarczone tak dużym błędem, iż prawdopodobnie jednak diagnoza jest fałszywa (zobacz rozdział *Medycyna*). Twierdzenie Bayesa to we wszystkich dziedzinach nauki potężne narzędzie analizowania i rozumienia tego, co naprawdę wiemy.

Bayesowski sposób myślenia pokazuje się z najlepszej strony dopiero wtedy, kiedy bierze się pod uwagę więcej niż jedną hipotezę – na przykład gdy stawia się diagnozę pacjentowi na podstawie objawów[14*] albo ustala pozycję samochodu autonomicznego na podstawie informacji dostarczanych przez czujniki. W teorii każda możliwa choroba czy współrzędna na mapie może odpowiadać rzeczywistości – zadanie polega więc na wybraniu na podstawie posiadanych informacji tej najbardziej prawdopodobnej.

Okazuje się, że problem zlokalizowania samochodu bez kierowcy jest bardzo zbliżony do problemu spędzającego sen z powiek Thomasowi Bayesowi, prezbiteriańskiemu duchownemu i utalentowanemu matematykowi, którego imieniem nazwano później słynne twierdzenie. W połowie XVIII wieku napisał on esej zawierający opis gry, którą wymyślił w celu zilustrowania owego zagadnienia. Brzmiał on mniej więcej tak[32]: Wyobraź sobie, że siedzisz plecami do kwadratowego stołu. Nie widzisz więc rzuconej przeze mnie na stół czerwonej kulki – twoim zadaniem jest domyślić się, gdzie się zatrzymała. Nie będzie to proste: nie masz żadnych informacji i nie wiesz, w której części stołu obecnie się znajduje.

A zatem żeby ci pomóc, rzucam na stół drugą kulkę, innego koloru. Twoje zadanie się nie zmieniło – masz ustalić położenie pierwszej, czerwonej kulki – tym razem jednak powiem ci, gdzie zatrzymała się druga kulka w stosunku do pierwszej: czy jest przed nią, czy za nią, na prawo czy na lewo od niej. Dzięki temu będziesz mógł uaktualnić swoje domysły.

Potem powtarzamy procedurę – rzucam na stół trzecią, czwartą, piątą kulkę i za każdym razem informuję cię, gdzie się zatrzymała w stosunku do pierwszej, czerwonej kulki – tej, której położenie cały czas próbujesz odgadnąć.

Im więcej kulek rzucę i im więcej podam ci informacji, tym dokładniejszy obraz stołu i umiejscowienia na nim czerwonej kulki powinien rysować się w twoich myślach. Nigdy nie będziesz miał całkowitej pewności, gdzie się dokładnie znajduje, ale możesz uaktualniać swoje dotychczasowe przekonanie po każdym rzucie, aż w końcu będziesz miał odpowiedź, której będziesz pewny.

W pewnym sensie rzeczywiste położenie samochodu autonomicznego odpowiada położeniu czerwonej kulki na stole. Zamiast osoby siedzącej plecami do stołu mamy algorytm, który stara się ustalić dokładne położenie pojazdu w danej chwili, a zamiast pozostałych kulek rzucanych na stół mamy różne źródła danych: GPS, czujniki bezwładności i tak dalej. Żadne z nich nie powie algorytmowi, gdzie jest samochód, każde jednak doda nową informację, której algorytm użyje do uaktualnienia swoich domysłów. Sztuczka ta nosi nazwę wnioskowania probabilistycznego, a polega na wykorzystaniu danych (i Bayesa) do wydedukowania rzeczywistego położenia przedmiotu. W odpowiednim opakowaniu będzie to kolejna odmiana algorytmu uczenia maszynowego.

Na przełomie tysiącleci inżynierowie mieli już tyle praktyki z pociskami manewrującymi, statkami kosmicznymi i samolotami, że umieli właściwie podejść do problemu pozycjonowania. Nauczenie samochodu bez kierowcy, żeby udzielał poprawnych odpowiedzi na pytanie: „Gdzie jestem?", nadal nie było łatwe, niemniej po uwzględnieniu myślenia bayesowskiego zdawało się przynajmniej celem osiągalnym.

Na drodze od cmentarzyska robotów na wyścigu Grand Challenge w 2004 roku do niewiarygodnego triumfu techniki podczas następnego wyścigu w 2005 roku – kiedy pięć różnych samochodów zdołało przejechać

ponad 200 kilometrów bez jakiejkolwiek pomocy człowieka – wiele z tych największych postępów zawdzięczamy Bayesowi. Algorytmy oparte na jego ideach pomogły znaleźć odpowiedzi na pytania, bez których pojazdy nie mogłyby jeździć: „Co mnie otacza?" i „Co mam robić?"[15*].

A więc – czy twój samochód autonomiczny ma przejechać przechodnia, żeby cię ratować?

Poświęćmy teraz chwilę na zastanowienie się nad drugim z powyższych pytań. Bo właśnie na ten temat jesienią roku 2016, w cichym kąciku bardzo gwarnego holu wystawy Paris Auto Show, wypowiedział się w raczej niespotykany sposób Christoph von Hugo, dyrektor wydziału systemów wspomagania kierowcy i aktywnego bezpieczeństwa. Na pytanie, co może zrobić autonomiczny mercedes podczas wypadku, odpowiedział tak:

„Jeżeli wiesz, że możesz uratować przynajmniej jedną osobę, uratuj chociaż tę jedną"[33].

W sumie to logiczne, można by pomyśleć.

Z tym że pana Hugo nie pytano o konkretny wypadek. Jego rozmówca chciał poznać zdanie dyrektora na temat znanego eksperymentu myślowego z lat sześćdziesiątych ubiegłego wieku, w którym omawiano bardzo szczególny wypadek drogowy i konieczność wybrania między dwoma złymi rozwiązaniami. Był to tak zwany dylemat wagonika, bo tak nazywano niedający się zatrzymać pojazd użyty w pierwotnym opisie eksperymentu. Wersja z udziałem samochodu bez kierowcy wyglądałaby następująco:

Wyobraź sobie, że za kilka lat będziesz pasażerem pojazdu autonomicznego, który pędzi beztrosko przez ulice miasta. Na najbliższym skrzyżowaniu zapala się czerwone światło, lecz błąd w układzie samochodu uniemożliwia jego zatrzymanie. Zderzenie jest nieuniknione, ale auto ma wybór: zjechać z drogi i uderzyć w betonową ścianę, zabijając wszystkich pasażerów, czy raczej jechać dalej prosto i przejechać przechodniów, którzy właśnie weszli na jezdnię? Jak powinien być zaprogramowany samochód? Na jakiej podstawie zdecydujesz, kto powinien umrzeć?

Nie wątpię, że masz na ten temat wyrobioną opinię. Być może uważasz, że samochód powinien po prostu ratować tyle osób, ile to możliwe. A może jesteś zdania, że piąte przykazanie powinno stać ponad wszelkimi kalkulacjami i człowiek siedzący w jeżdżącej maszynie powinien ponieść konsekwencje swoich czynów?[16*]

Hugo jasno wyraził stanowisko firmy Mercedes. „Uratuj człowieka, który siedzi w samochodzie". Potem jeszcze dodał: „Jeżeli wiesz na pewno tylko tyle, że możesz zapobiec śmierci jednego człowieka, to właśnie powinien być twój priorytet".

Kilka dni po tym wywiadzie internet już pękał w szwach od artykułów krytykujących stanowisko Mercedesa. „Ich samochody będą się zachowywały dokładnie tak samo jak stereotypowy uprzywilejowany Europejczyk w swoim luksusowym samochodzie", napisał jeden z autorów[34]. W ankiecie opublikowanej w czasopiśmie „Science" tego samego lata[35] 76 procent respondentów uznało, że byłoby bardziej etycznie, gdyby samochód autonomiczny starał się uratować jak najwięcej ludzi – za cenę życia swoich pasażerów. Wyglądało więc na to, że Mercedes opowiedział się po stronie zdecydowanej mniejszości opinii publicznej.

Ale czy rzeczywiście? Kiedy autorzy tego samego studium spytali respondentów, czy k u p i l i b y samochód, który w szczególnych okolicznościach mógłby ich zamordować, nagle się okazało, że mało kto był gotów poświęcać siebie dla innych.

Tego rodzaju dylematy dzielą opinię publiczną, i to nie tylko z powodu różnic w wyborze właściwej odpowiedzi. Eksperyment myślowy z wagonikiem pozostaje od lat faworytem dziennikarzy i komentatorów nowoczesnej techniki, lecz gdy tylko o nim wspominałam, eksperci od pojazdów autonomicznych zaczynali wzdychać i przewracać oczami. Ja mam do tego eksperymentu wielką słabość; jego prostota zmusza nas do dostrzeżenia poważnego problemu związanego z pojazdami autonomicznymi: jak się nam podoba to, że algorytm będzie podejmował decyzje etyczne, od których będzie zależało życie nasze i innych ludzi? Podstawą tej nowej dziedziny techniki – podobnie jak prawie wszystkich algorytmów – są zagadnienia władzy, kontroli i delegowania odpowiedzialności. Czy możemy liczyć na to, że ta technika dopasuje się

do nas, a nie na odwrót? Rozumiem jednak także raczej obojętny stosunek zwolenników transportu autonomicznego do tego eksperymentu – ci ludzie lepiej niż ktokolwiek inny wiedzą, jak bardzo daleko jesteśmy jeszcze od rzeczywistości, w której musielibyśmy poważnie brać pod uwagę dylemat wagonika.

Łamanie zasad ruchu

Twierdzenie Bayesa i potęga rachunku prawdopodobieństwa stanowią główną siłę napędową innowacji w pojazdach autonomicznych od czasów pierwszego wyścigu DARPA Grand Challenge. Paul Newman jest profesorem robotyki na uniwersytecie w Oxfordzie i założycielem firmy Oxbotica, która konstruuje samochody autonomiczne i testuje je na brytyjskich ulicach. Zapytałam go o to, jak się sprawują jego najnowsze prototypy. Odpowiedział mi tak: „To miliony i jeszcze raz miliony wierszy kodu, ale całą rzecz ująłbym krótko jako wnioskowanie probabilistyczne”[36].

I choć wnioskowanie bayesowskie w dużej mierze wyjaśnia, jakim cudem samochód może jeździć bez kierowcy, pokazuje nam również, że osiągnięcie pełnej autonomii – całkowitej wolności od wszelkiej ingerencji człowieka – to nadzwyczaj twardy orzech do zgryzienia.

Paul Newman każe nam wyobrazić sobie dwa pojazdy zbliżające się do siebie z określoną prędkością – powiedzmy, że jadą w różnych kierunkach drogą jednopasmową po lekkim łuku. Człowiek w takiej sytuacji czuje się komfortowo, bo wie, że nadjeżdżający z przeciwka samochód będzie się trzymał swojego pasa ruchu i minie go w bezpiecznej odległości paru metrów. „Ale przez dłuższą chwilę faktycznie wygląda to tak, że dojdzie do zderzenia”, mówi Newman. Jak nauczyć autonomiczny samochód, żeby nie spanikował w podobnej sytuacji? Nie chcielibyśmy przecież, żeby zjechał nagle na pobocze dla uniknięcia wypadku, którego i tak by nie było. Ale też nie chcielibyśmy, żeby nasz samochód bez kierowcy zachowywał się beztrosko, bo zawsze może się zdarzyć sytuacja na granicy zderzenia czołowego. Pamiętajmy, że swoje decyzje o tym, co robić, pojazdy autonomiczne opierają wyłącznie na najbardziej prawdopodobnych

domysłach. Jak to zrobić, żeby te domysły były słuszne za każdym razem, bez wyjątku? Zdaniem Newmana „jest to szalenie trudny problem".

I choć problem ten od bardzo dawna spędza ekspertom sen z powiek, istnieje rozwiązanie. Cała sztuczka polega na tym, żeby zbudować model oparty na zachowaniu i reakcjach rozsądnego kierowcy. Niestety, modelu tego nie da się już zastosować do wielu specyficznych sytuacji, do których może dojść w ruchu drogowym.

Newman tłumaczy: „Trudne w prowadzeniu samochodu jest to wszystko, co z samym prowadzeniem nie ma nic wspólnego". Na przykład jak nauczyć algorytm, że na dźwięk pozytywki samochodu z lodami albo na widok dzieci z piłką na chodniku należy zdecydowanie zwiększyć ostrożność? Albo jak nauczyć go rozpoznawać podskoki kangura, z którymi (w czasie, kiedy pisałam tę książkę) pracująca również nad autonomicznymi samochodami firma Volvo nie mogła sobie poradzić?[37] To niewielkie zmartwienie, jeśli mieszkasz w Europie, ale zanim samochody autonomiczne zostaną dopuszczone do ruchu w Australii, będą musiały się tych kangurów „nauczyć".

Jeszcze trudniej jest nauczyć samochód, że czasem trzeba złamać przepisy ruchu drogowego. Wyobraź sobie, że twój autonomiczny pojazd stoi na czerwonym świetle i nagle ktoś wybiega ci przed maskę i żywo gestykulując, każe ci zjechać na bok. Albo karetka na sygnale nie może cię wyprzedzić w wąskiej uliczce, więc musisz wjechać na chodnik, żeby ją przepuścić. Albo tankowiec wbił się w wybrzeże i przeciął ci drogę, więc musisz uciekać najszybciej i najdalej, jak tylko się da.

„Takich sytuacji nie ma w kodeksie drogowym", słusznie zauważa Newman. A mimo to pojazd naprawdę autonomiczny musi wiedzieć, jak sobie z nimi radzić, jeżeli ma jeździć bez ingerencji człowieka. Nawet w przypadkach zagrożenia życia.

Nie znaczy to wcale, że są to problemy nie do przezwyciężenia. „Nie wierzę, żeby istniał taki poziom inteligencji, do jakiego nie bylibyśmy w stanie przyuczyć maszyny – powiedział mi Newman. – Pytanie tylko, kiedy to nastąpi".

Niestety, odpowiedź na to pytanie brzmi: prawdopodobnie nie za szybko. Spełnienie ludzkiego marzenia o s a m o c h o d z i e, pojeździe

prawdziwie samojezdnym, może być znacznie odleglejsze w czasie, niż nam się dziś wydaje.

Jest bowiem jeszcze jeden poziom trudności, który musimy pokonać, żeby wcielić w życie tę fantastycznonaukową wizję autonomicznego samochodu bez kierowcy i kierownicy, dla którego nie ma rzeczy niemożliwych. Pojazd w pełni autonomiczny będzie musiał również poradzić sobie z wcale nie łatwym problemem człowieka.

Jack Stilgoe, socjolog z University College London i ekspert od wpływu techniki na życie społeczne, wyjaśnia, w czym rzecz: „Ludzie to złośliwcy. To podmioty aktywne, a nie bierne elementy scenografii”[38].

Wyobraźmy sobie przez chwilę świat, w którym naprawdę istnieją doskonale autonomiczne pojazdy. Zasadą numer jeden zakodowaną w ich komputerze pokładowym jest unikanie kolizji za wszelką cenę. A to zmienia dynamikę ruchu drogowego. Kiedy ktoś wejdzie przed maskę takiego pojazdu – będzie on musiał się zatrzymać. Jeżeli zajedzie się mu drogę na skrzyżowaniu – będzie musiał ustąpić.

Jedna z uczestniczek grupy fokusowej zorganizowanej w 2016 roku w London School of Economics powiedziała: „Ludzie od razu zaczną wymuszać na nich pierwszeństwo. Autonomiczne pojazdy będą się zatrzymywać, a oni będą je po prostu objeżdżać”. W swobodnym tłumaczeniu: naiwne samochody autonomiczne będą wykorzystywane przez cwaniaków.

Stilgoe zgadza się z tą opinią: „Do tej pory stosunkowo słabi uczestnicy ruchu, na przykład rowerzyści, mogą zacząć jeździć bardzo wolno przed takimi samochodami, wiedząc, że nigdy nie wzbudzą swoim zachowaniem agresywnej reakcji”.

Niewykluczone, że rozwiązanie tego problemu może oznaczać wprowadzenie surowszych zasad dla rowerzystów czy pieszych nadużywających przysługujących im praw. Już to przerabialiśmy, na przykład z nieprawidłowym przechodzeniem przez jezdnię. Może to oznaczać również wypchnięcie większości uczestników ruchu z dróg, zupełnie jak kiedyś, tuż po pojawieniu się pojazdu z silnikiem spalinowym. Z tej samej przyczyny nie widzimy już na autostradach rowerów, zaprzęgów konnych ani pieszych.

Jeżeli chcemy mieć w pełni autonomiczne samochody, niemal na pewno będziemy musieli zrobić coś podobnego raz jeszcze i ograniczyć liczbę agresywnych kierowców, samochodów do sprzedaży lodów, dzieciaków grających na ulicy w piłkę, robót drogowych, niefrasobliwych przechodniów, pojazdów uprzywilejowanych, rowerzystów, wózków inwalidzkich i w ogóle wszystkiego, co tak bardzo komplikuje problem autonomii. No, ale nie tak przecież miało to wyglądać; zupełnie inną wizję próbuje się nam sprzedawać.

„Cała retoryka autonomii i transportu opiera się na tym, żeby n i e z m i e n i a ć naszego świata – mówi Stilgoe. – Wszystko ma zostać tak, jak jest, i to robot ma najpierw dorównać człowiekowi, a potem go prześcignąć w sztuce poruszania się po tym niezmienionym świecie. Dla mnie to idiotyzm".

Ale chwileczkę. Czy ten problem nie został już dawno rozwiązany? Czy Waymo, autonomiczny samochód Google'a, nie przejechał już milionów kilometrów? Czy całkowicie autonomiczne samochody tej marki (no, prawie całkowicie autonomiczne) nie jeżdżą po ulicach Phoenix, stolicy Arizony?

No tak. To prawda. Tylko że kilometr kilometrowi nierówny. Większość z nich jest tak prosta, że można je przejechać z zamkniętymi oczami. Inne znów są dużo bardziej wymagające. W czasie, kiedy piszę tę książkę, samochody Waymo nie mogą jeszcze jeździć wszędzie, ponieważ zostały poddane „geograniczeniu" (ang. *geofencing*): mogą się poruszać w niewielkim, z góry wyznaczonym obszarze. Podobnie będzie z autonomicznymi samochodami, które Daimler i Ford chcą wypuścić na drogi odpowiednio w 2020 i 2021 roku. Będą one używane w usługach przewozowych w ściśle określonej strefie, dzięki czemu problem autonomii stanie się zdecydowanie prostszy do rozwiązania.

Paul Newman uważa, że tak właśnie będzie wyglądać najbardziej prawdopodobna przyszłość autonomicznych pojazdów: „Będą się poruszać w doskonale znanym sobie obszarze, gdzie ich właściciele będą absolutnie pewni, że nie przysporzą im żadnych problemów. Gdzieś w mieście, z dala od miejsc, gdzie drogi są niestandardowe albo gdzie przed maskę może nagle wyjść krowa. Może będą jeździły w określonych porach dnia i przy określonej pogodzie. Będą używane w usługach transportowych".

To zdecydowanie nie to samo co pełna autonomia. Tak Jack Stilgoe wyobraża sobie ów konieczny kompromis: „To, co wygląda jak system autonomiczny, jest w rzeczywistości systemem, dla którego na świat nałożono pewne ograniczenia, żeby mu nadać pozory autonomii".

Wizja, w którą nam każą wierzyć, to zwykła sztuczka z cieniami. Obraz, który zdaje się obiecywać każdemu luksus osobistego szofera, a z bliska okazuje się busikiem komunikacji gminnej.

Kropkę nad i niech postawi jedno z największych amerykańskich czasopism motoryzacyjnych „Car and Driver": „Żaden producent samochodów nie spodziewa się w najbliższej przyszłości – ani nawet w ciągu najbliższych kilkudziesięciu lat – urzeczywistnienia utopijnej wizji bezwypadkowego świata z ulicami rojącymi się od autonomicznych pojazdów. Wszyscy oni chcą raczej zrobić wrażenie na Wall Street i stymulować wyobraźnię klientów, którzy są coraz mniej zainteresowani samodzielnym prowadzeniem samochodu. Tak naprawdę jednak liczą tylko na to, że będą mogli dalej sprzedawać mnóstwo aut wyposażonych w najnowocześniejsze, bardzo skomplikowane systemy wsparcia kierowcy"[39].

Co zatem z owymi systemami wsparcia kierowcy? Ostatecznie samochód autonomiczny nie musi być problemem zerojedynkowym.

Technikę autonomiczną w motoryzacji dzieli się na sześć kategorii: od poziomu 0 – brak jakiejkolwiek automatyki – po poziom 5, czyli w pełni autonomiczna fantastyka naukowa. Pomiędzy tymi skrajnościami mamy różne formy automatyki: od tempomatu (poziom 1) po pojazdy autonomiczne z „geograniczeniem" (poziom 4). Potocznie mówi się o „wyłączeniu nogi" (poziom 1), „wyłączeniu rąk" (poziom 2), „wyłączeniu oczu" (poziom 3) i „wyłączeniu mózgu" (poziom 4).

Tak więc wiele wskazuje na to, że choć poziom 5 jest jeszcze daleko poza naszym zasięgiem, a poziom 4 nie będzie wcale taki, jak nam obiecują, czeka nas po drodze cała masa nowych odmian automatyki. Powolne podnoszenie poziomu wsparcia kierowcy w naszych prywatnych samochodach chyba nie może zaszkodzić? Co powiecie na quasi-autonomiczne samochody z normalnymi pedałami i tradycyjną kierownicą, za którą można usiąść i w razie konieczności interweniować?

To nam chyba wystarczy do czasu, kiedy technika zrobi kolejny skok do przodu?

Niestety, sprawa nie jest aż taka prosta. Bo ta historia ma jeszcze jeden wymiar, w którym roi się od nowych problemów nie do uniknięcia we wszelkich pojazdach poniżej poziomu 5, czyli autonomii absolutnej, wolnej od jakiejkolwiek ingerencji człowieka.

Dziecko firmy

Wśród pilotów Air France Pierre-Cédric Bonin był znany jako „dziecko firmy”[40]. Zaczął pracować jako pilot ze skromniutkim nalotem kilkuset godzin, w wieku zaledwie 26 lat, i dorastał razem z francuskimi airbusami. Kiedy sześć lat później wchodził na pokład samolotu obsługującego feralny lot AF447, miał już wylataną imponującą liczbę 2936 godzin, ale nadal był najmniej doświadczonym z trzech pilotów załogi[41].

Niemniej to właśnie Bonin siedział za sterami, kiedy 31 maja 2009 roku samolot oderwał się od pasa startowego międzynarodowego lotniska Rio de Janeiro-Galeão i wziął kurs na Paryż[42].

Był to airbus A330, jeden z najbardziej zaawansowanych technicznie samolotów, jakie kiedykolwiek wyprodukowano dla lotnictwa cywilnego. Autopilot w tych maszynach jest praktycznie bezobsługowy i może sterować samolotem przez cały czas z wyjątkiem startu i lądowania. I nawet wówczas kiedy za sterami siedzi człowiek, szereg zabezpieczeń minimalizuje ryzyko popełnienia błędu.

Skonstruowanie zautomatyzowanego systemu zdolnego bezpiecznie rozwiązać dosłownie każdy problem, który potrafią przewidzieć jego projektanci, tworzy jednak niejawne zagrożenie. Jeśli rola pilota ogranicza się do przejmowania sterów wyłącznie w sytuacjach wyjątkowych, człowiek przestaje ćwiczyć umiejętności i nawyki niezbędne do bezpiecznego pilotowania samolotu bez pomocy komputerów. Nie nabywa więc doświadczenia, z którego mógłby skorzystać w razie wystąpienia nieoczekiwanego zagrożenia.

Taki był właśnie przypadek lotu AF447. Mimo że Bonin nalatał tysiące godzin w kokpicie airbusa, jego realne doświadczenie w samodzielnym

pilotowaniu A330 było znikome. Jego praca sprowadzała się zasadniczo do monitorowania systemu automatycznego. Kiedy więc akurat tego dnia autopilot się wyłączył, Bonin po prostu nie umiał bezpiecznie sterować maszyną[43].

Kłopoty zaczęły się od tego, że czujniki prędkości zaczęły zamarzać; autopilot przestał odbierać informacje o prędkości lotu, włączył więc alarm w kokpicie i przekazał stery człowiekowi. To jeszcze nie było powodem do zmartwienia, kiedy jednak później samolot wpadł w strefę niewielkich turbulencji, niedoświadczony Bonin zareagował zbyt gwałtownie: przy nieznacznym przechyle maszyny w prawo chwycił drążek sterowy i odchylił go w lewo. Najgorsze jednak, że jednocześnie pociągnął go do siebie, rozpoczynając gwałtowny manewr wznoszenia[44].

Bonin dalej kurczowo trzymał drążek, samolot wzbijał się w rzednącej atmosferze, aż w końcu dziób maszyny znalazł się tak wysoko, że skrzydła zamiast ciąć powietrze, zaczęły działać jak hamulce; siła nośna drastycznie zmalała i wychylony w górę samolot zaczął spadać z nieba jak kamień.

W kokpicie rozległ się kolejny alarm. Wybudzony z drzemki kapitan wybiegł z kabiny odpoczynku. AF447 spadał do oceanu, tracąc wysokość w tempie 3200 metrów na minutę.

Czujniki prędkości tymczasem odmarzły, nie było żadnej usterki mechanicznej, do oceanu było jeszcze daleko i sytuację można było opanować. Bonin i drugi pilot w kokpicie mogli bez trudu uratować wszystkich pasażerów i całą załogę w 10–15 sekund – wystarczyło wychylić drążek sterowy w przód, żeby opuścić nos maszyny i przywrócić swobodny przepływ powietrza nad i pod skrzydłami[45].

Spanikowany Bonin nadal jednak ciągnął drążek do siebie – ani on, ani jego kolega w kokpicie nie zdawali sobie sprawy, że to właśnie jest powodem spadania samolotu. Uciekały cenne sekundy. Kapitan zasugerował wyrównanie skrzydeł. Po krótkiej wymianie opinii, czy już się wznoszą, czy jeszcze spadają, na wysokości 2560 metrów nad poziomem morza stery przejął drugi pilot[46].

„W górę! W górę! W górę!", słychać krzyk drugiego pilota.

„Przecież cały czas ciągnąłem drążek do siebie!", odkrzyknął Bonin.

Dopiero wtedy kapitan zrozumiał, co się dzieje – że od trzech minut spadają w stanie przeciągnięcia – i kazał im natychmiast opuścić dziób samolotu. Na to było już jednak za późno, znajdowali się tragicznie nisko. Bonin krzyknął: „Kurwa, rozbijemy się! To niemożliwe!”[47]. Kilka chwil później samolot zderzył się z wodą. Zginęli wszyscy, 228 osób.

Paradoksy automatyki

W 1983 roku, dwadzieścia sześć lat przed katastrofą samolotu Air France, psycholożka Lisanne Bainbridge napisała głośny esej o niejawnych zagrożeniach wynikających ze zbytniego polegania na systemach zautomatyzowanych[48]. Skonstruuj maszynę poprawiającą sprawność człowieka – dowodziła autorka – a paradoksalnie, skończy się na tym, że ograniczysz ludzkie możliwości.

Dziś już każdy z nas zdążył się o tym przekonać na własnej skórze: nie pamiętamy już numerów telefonów, wielu z nas z trudem odczytuje własne pismo i nie potrafi już nigdzie dojechać bez nawigacji. Kiedy technika robi wszystko za nas, nie mamy kiedy ćwiczyć naszych umiejętności.

Są obawy, że taki scenariusz może się powtórzyć także z samochodami autonomicznymi – gdzie stawka jest o wiele większa niż przy odczytywaniu pisma odręcznego. Zanim dojdziemy do pełnej autonomii, nasze samochody będą od czasu do czasu nieoczekiwanie przekazywać nam kontrolę nad pojazdem. Czy instynkt nas nie zawiedzie, czy będziemy jeszcze pamiętać, jak się należy zachować? A co z nastoletnimi kierowcami przyszłości – czy kiedykolwiek będą mieli możliwość zdobycia umiejętności niezbędnych do zapanowania nad samochodem?

Ale choćby nawet wszyscy kierowcy zdołali zachować te niezbędne umiejętności[17*] (nie będziemy teraz precyzować, co przez te „umiejętności” należy rozumieć), pozostanie nam do rozwiązania jeszcze jeden problem. Bardzo wiele będzie zależeć od tego, czego zażąda od człowieka autopilot, zanim się wyłączy. Możliwości są tylko dwie i – jak podkreśla Bainbridge – żadna nie może napawać optymizmem.

Samochód z poziomu 2 automatyki („wyłączone ręce") będzie oczekiwał od kierowcy bacznego obserwowania drogi przez cały czas[49]. Komputer w takim pojeździe nie jest jeszcze tak dobry, by można mu było wierzyć, i dlatego będzie wymagał stałego nadzoru. Jeden z autorów czasopisma „Wired" opisał kiedyś ten poziom automatyki następująco: „To tak, jakbyś chciał nakłonić raczkujące dziecko, żeby ci pomogło pozmywać naczynia"[50].

Autopilot Tesli to dobry przykład takiego podejścia[51]. Obecnie przypomina on raczej skomplikowany tempomat – może kierować, hamować i przyspieszać na autostradzie, wymaga jednak od kierowcy stałej czujności i gotowości do interwencji w każdej chwili. Aby mieć pewność, że rzeczywiście uważasz, samochód włącza alarm, jeśli na zbyt długą chwilę oderwiesz ręce od kierownicy.

Bainbridge napisała, że to się nie może skończyć dobrze. Wymaganie od człowieka stałej czujności jest po prostu nierealne: „Nawet osoba bardzo zmotywowana nie jest w stanie skupić wzroku na źródle informacji, z którym dzieje się bardzo niewiele, przez czas dłuższy niż jakieś pół godziny"[52].

Są pewne dowody świadczące o tym, że ludzie rzeczywiście mają spore problemy z zastosowaniem się do wymogu zachowania czujności podczas jazdy z autopilotem Tesli. Joshua Brown, który zginął za kierownicą swojej tesli w 2016 roku, jechał w trybie automatycznej jazdy od prawie 38 minut, kiedy jego samochód uderzył w ciężarówkę przecinającą jego pas ruchu. Dochodzenie przeprowadzone przez amerykański Urząd do Spraw Bezpieczeństwa Drogowego wykazało, że w chwili wypadku Brown nie patrzył na drogę[53]. Mimo że o wypadku pisano na pierwszych stronach gazet na całym świecie, niereformowalni youtuberzy nadal zamieszczali w sieci entuzjastyczne filmiki z receptami na oszukanie autopilota i pozorowanie czujności. Podobno wystarczy przykleić taśmą do kierownicy puszkę Red Bulla[54] albo wcisnąć w kierownicę pomarańczę[55], żeby samochód przestał włączać wkurzające alarmy przypominające człowiekowi o jego powinnościach.

Inne programy borykają się z tym samym problemem. Mimo że autonomiczny samochód Ubera wymaga interwencji człowieka co mniej

więcej 21 kilometrów[56], skłonienie kierowców do zachowania czujności nie wydaje się możliwe. Osiemnastego marca 2018 roku autonomiczny pojazd Ubera potrącił śmiertelnie kobietę przechodzącą przez jezdnię. Film z wnętrza pojazdu wykazał, że siedzący za kierownicą „ludzki nadzorca" w chwilach poprzedzających potrącenie nie patrzył na drogę[57].

Problem okazuje się poważny, lecz jest pewna nadzieja. Producenci samochodów mogliby pogodzić się z tym, że jesteśmy tylko ludźmi i że nigdy nie będziemy w stanie skupić się dłuższy czas. Jakkolwiek by było, możliwość poczytania książki podczas długiej jazdy to jeden z d u ż y c h p l u s ó w samochodów autonomicznych. To właśnie kluczowa różnica między poziomem 2 – „wyłączone ręce", a poziomem 3 – „wyłączone oczy".

Poziom 3 stanowi z pewnością znacznie poważniejsze wyzwanie techniczne niż poziom 2, ale niektórzy producenci już zaczęli konstruować samochody, biorąc poprawkę na naszą nieuważność. Za przykład może nam posłużyć asystent do jazdy w korkach firmy Audi[58], który jest w stanie całkowicie przejąć kierowanie autem, kiedy stoisz w wolno poruszającym się ciągu samochodów – możesz się wtedy wygodnie oprzeć w fotelu i dać się wieźć przed siebie. Musisz być tylko gotowy do awaryjnej interwencji[18*].

Jest powód, dla którego firma Audi ograniczyła funkcjonowanie swojego systemu do powolnego ruchu na drogach szybkiego ruchu. W korku na takiej drodze ryzyko katastrofy jest mniejsze. Kierowca, który nie patrzy na drogę podczas jazdy, będzie miał bardzo mało czasu na ocenę sytuacji i podjęcie właściwej decyzji. Wyobraź sobie, że siedzisz w samochodzie autonomicznym i nagle słyszysz alarm – podnosisz głowę znad książki i widzisz, że ciężarówka przed tobą gubi ładunek, w który zaraz wjedziesz. Będziesz musiał błyskawicznie przetworzyć wszystkie informacje z otoczenia: motocykl na prawym pasie, przed tobą gwałtownie hamująca ciężarówka, samochód w martwym polu po lewej. Nie będziesz miał pojęcia o sytuacji na drodze akurat w tym momencie, kiedy powinieneś wiedzieć o niej wszystko; dodaj do tego brak praktyki, a sam stwierdzisz, że jesteś po prostu nieprzygotowany do poradzenia sobie z sytuacją wymagającą od kierowcy najwyższych umiejętności.

Obawy zostały potwierdzone w eksperymentach na symulatorach pojazdów autonomicznych. Jedno z badań, w których pozwolono ludziom czytać albo grać na telefonie, podczas gdy samochód prowadził się sam, wykazało, że po usłyszeniu alarmu badani potrzebowali nawet 40 sekund, żeby odzyskać pełną kontrolę nad pojazdem[59]. Właśnie coś takiego wydarzyło się podczas lotu Air France 447. Kapitan Dubois, który mógł bez większego trudu uratować samolot, stracił około minuty, zanim się zorientował, co się dzieje i na czym polega problem[60].

Paradoksalnie, im lepszą mamy technikę autonomizacji pojazdów, w tym większe brniemy kłopoty. Z leniwym autopilotem, który co kwadrans włącza alarm, kierowca będzie musiał zachowywać bezustanną czujność i regularnie ćwiczyć swoje umiejętności. Strzeżmy się natomiast wszystkich eleganckich, bardzo złożonych systemów automatyki, które potrafią sobie poradzić w niemal każdych warunkach.

Zdaniem szefa instytutu badawczego Toyoty Gilla Pratta w najgorszym scenariuszu samochód będzie wymagał awaryjnej interwencji człowieka raz na 300 tysięcy kilometrów. Ponieważ przeciętny kierowca zmienia auto po przejechaniu 150 tysięcy kilometrów, może się zdarzyć, że dopiero w drugim czy trzecim samochodzie autonomicznym spotka się z sytuacją, kiedy komputer nagle przekaże mu prowadzenie pojazdu – a że od lat nie prowadził sam i nie był w niebezpiecznej sytuacji, nie będzie zupełnie przygotowany na coś takiego[61].

Wielkie oczekiwania

A jednak mamy ważne powody, żeby dążyć do autonomicznej przyszłości. Plusy ciągle przeważają nad minusami. Wypadki komunikacyjne to wciąż jedna z głównych przyczyn dających się uniknąć śmierci na świecie. Jeśli technika automatyzacji mogłaby choć trochę ograniczyć liczbę ofiar śmiertelnych na drogach, można by śmiało twierdzić, że zaniechanie w tym względzie byłoby nieetyczne.

Nie brakuje również innych korzyści: nawet proste systemy wsparcia kierowcy mogą ograniczyć zużycie paliwa[62] i poprawić przepustowość dróg[63]. No i powiedzmy to sobie szczerze: myśl o puszczeniu kierownicy

przy prędkości 120 kilometrów na godzinę, nawet na krótką chwilę, jest po prostu… niesamowita.

Wróćmy jednak do ostrzeżeń Bainbridge – w istocie wskazują one na problematyczny sposób opisywania współczesnej techniki pojazdów autonomicznych.

Weźmy Teslę, jednego z pierwszych producentów, którzy wprowadzili na rynek autopilota. Mało kto zaprzecza, że system wprowadzony przez ten koncern miał ogólny pozytywny wpływ – kierowanie teslami jest bezpieczniejsze dla ich użytkowników, a kto nie jest przekonany, niech odszuka w internecie filmiki pokazujące działania funkcji ostrzeżenie przed możliwą kolizją (ang. *Forward Collision Warning*), która dostrzega ryzyko wypadku, nim zrobi to kierowca, włącza alarm i zapobiega zderzeniu[64].

Istnieje jednak subtelna różnica między tym, co te samochody potrafią – z wyposażeniem będącym zasadniczo wyrafinowanym połączeniem przedniego czujnika parkowania z inteligentnym tempomatem – a językiem używanym do opisu tych możliwości. Na przykład w październiku 2016 roku Tesla wydała oświadczenie, w którym można przeczytać, że „wszystkie produkowane obecnie samochody naszej firmy mają pełne oprzyrządowanie potrzebne do samosterowania”[19*]. Jak donosi portal Verge, założyciel firmy Tesla Elon Musk dodał: „Aktualizacja do pełnej autonomii będzie od teraz standardem we wszystkich naszych pojazdach”[65]. Nietrudno zauważyć, że wyrażenie „pełna autonomia” niezbyt pasuje do ostrzeżenia, które użytkownicy muszą zaakceptować, zanim uruchomią instalowanego w teslach autopilota: „Musisz panować nad swoim pojazdem i być za niego odpowiedzialny”[66].

Oczekiwania są ważne. Możesz się z tym nie zgodzić, ale moim zdaniem to, że ludzie wciskają pomarańcze w kierownicę – albo jeszcze gorzej, o czym się przekonałam, przeszukując mroczniejsze zakamarki sieci: produkują i sprzedają urządzenia, które „umożliwiają pierwszym użytkownikom nowego systemu Tesli jazdę z ograniczonym albo wyłączonym alarmem autopilota”[20*] – jest nieuniknioną konsekwencją faktu, że szanowana marka używa języka, który wprowadza klientów w błąd.

Oczywiście Tesla nie jest w przemyśle samochodowym jedyną firmą, która dopuszcza się podobnych praktyk – każde przedsiębiorstwo na świecie sprzedaje swoje produkty, odwołując się do naszych marzeń. A jednak uważam, że jest różnica między kupowaniem perfum w nadziei, że stanę się dzięki nim bardziej atrakcyjna, a kupowaniem samochodu z przekonaniem, że jego „pełna autonomiczność" zapewni mi bezpieczeństwo.

Abstrahując od strategii marketingowych, nie mogę przestać myśleć, że nasze wyobrażenia o samochodzie bez kierowcy poszły w zupełnie niewłaściwym kierunku.

Wiemy już, że człowiek świetnie sobie radzi z rozumieniem niuansów, analizowaniem kontekstu, wyciąganiem wniosków z własnych doświadczeń i rozpoznawaniem prawidłowości. Bardzo źle natomiast wypada w testach czujności, precyzji, konsekwencji i pełnej świadomości własnego otoczenia. Krótko mówiąc, nasz zestaw mocnych i słabych stron to dokładne przeciwieństwo zestawu mocnych i słabych stron algorytmów.

A więc dlaczego nie pójść za światem medycyny z jego programami wyszukiwania guzów i nie pozwolić algorytmom dopełniać człowieka, z korzyścią dla obu stron? Dlaczego nie odwrócić kota ogonem i dopóki nie dojdziemy do pełnej autonomii – nie dążyć do stworzenia systemu prowadzenia, w którym komputer wspierałby kierowcę, a nie odwrotnie? Stwórzmy system bezpieczeństwa, taki jak ABS czy kontrola trakcji, który cierpliwie monitorując drogę, będzie bezustannie czuwał, w razie gdyby kierowca przegapił niebezpieczeństwo. Byłby to nie szofer, ale raczej stróż.

Ta idea przyświeca konstruktorom z instytutu badawczego Toyoty. Pracują nad dwoma trybami pracy autopilota: pierwszy to tryb „szofera", który – podobnie jak autopilot Audi – może przejmować kierownicę w korkach, oraz tryb drugi: „stróża", który pracuje w tle, kiedy człowiek kieruje – to właśnie taki system bezpieczeństwa[67] zmniejszający ryzyko wypadku, gdyby nagle na drodze pojawiło się niedostrzeżone przez kierowcę zagrożenie ruchu.

Volvo już obrało podobny kierunek rozwoju. W powszechnej opinii to właśnie dzięki szwedzkiemu systemowi autonomicznego hamowania awaryjnego (ang. *Autonomous Emergency Braking System*), który automatycznie zmniejsza prędkość samochodu, gdy ten podjeżdża za

blisko pojazdu znajdującego się przed nim, volvo XC90 ma tak imponujące statystyki bezpieczeństwa: odkąd auto to zostało wprowadzone do sprzedaży w Wielkiej Brytanii w 2002 roku, sprzedano ponad 50 tysięcy egzemplarzy, i ani jeden jego kierowca ani pasażer nie zginął w wypadku drogowym[68].

Jeśli chodzi o większość tak żywo dziś dyskutowanych rozwiązań techniki autonomizacji jazdy, będziemy musieli poczekać i zobaczyć, jak się to wszystko ułoży. To jednak wydaje się pewne: z biegiem czasu autonomiczne pojazdy uświadamiają nam parę rzeczy, które będą wykraczać daleko poza świat motoryzacji; nauczą nas mianowicie tego, że z trudem przejmujemy kontrolę w chaotycznych okolicznościach, oraz tego, że powinniśmy być bardziej realistyczni w naszych oczekiwaniach wobec możliwości algorytmów.

Wtedy będziemy musieli zmienić nasz sposób myślenia. Trzeba będzie odrzucić wyobrażenie, że autonomiczne samochody powinny działać bezawaryjnie zawsze i wszędzie, i pogodzić się z tym, że jeśli w najbliższej przyszłości awaria mechaniczna może być rzadkością, to awaria komputera z pewnością już nie.

Błędy są nieuniknione, a więc jeśli dalej chcemy podążać tą drogą, musimy po prostu pogodzić się z niepewnością. Dylematy, na jakie natrafimy w świecie pojazdów autonomicznych, zmuszą nas do określenia, jak dobre ma być nowe rozwiązanie, zanim uznamy, że można je testować na naszych ulicach. To ważne pytanie, które jest stawiane także w innych dziedzinach. Co właściwie znaczy „dobre”? Kiedy to „dobre” rozwiązanie jest wystarczająco dobre? Jeżeli skonstruujemy wadliwy algorytm, który mimo wszystko p o t r a f i coś obliczyć – czy powinniśmy mu na to pozwolić?

Przestępczość

W pewien ciepły lipcowy dzień 1995 roku w Leeds 22-letnia studentka spakowała książki, wyszła z biblioteki i ruszyła do swojego samochodu, który zostawiła na wielopoziomowym parkingu. Dzień minął jej na wprowadzaniu ostatnich poprawek do dysertacji, tak więc miała przed sobą letnie wakacje. Ale gdy usiadła za kierownicą i zaczęła się szykować do odjazdu, usłyszała, że ktoś biegnie w jej stronę. Zanim zdążyła zareagować, napastnik wsunął głowę przez otwarte okno i przystawił jej nóż do gardła. Kazał jej przesiąść się do tyłu, związał ją i skleił jej powieki klejem błyskawicznym, po czym sam usiadł za kierownicą i odjechał.

Po mrożącej krew w żyłach jeździe napastnik zatrzymał samochód przy trawiastym nabrzeżu. Zaraz potem kobieta usłyszała, jak mężczyzna opuszcza siedzenie i rozpina spodnie. Wiedziała, że będzie próbował ją zgwałcić, więc w ślepej furii podciągnęła kolana do piersi i kopnęła z całych sił, zmuszając napastnika, żeby się cofnął. Kopnęła jeszcze kilka razy, mężczyzna przeciął sobie palce trzymanym w ręku nożem i krew trysnęła na siedzenia. Napastnik dwukrotnie uderzył kobietę w twarz, ale potem nagle wysiadł z samochodu i uciekł. Studentka odetchnęła z ulgą. Dwie godziny po tym przerażającym porwaniu znaleziono ją przy Globe Road w Leeds; zrozpaczona kobieta szła po omacku ulicą, była w podartej spódnicy, potargana, twarz miała czerwoną po uderzeniach i sklejone powieki[1].

Tego rodzaju napaści seksualne na osoby nieznajome zdarzają się nadzwyczaj rzadko, ale kiedy już dojdzie do takiego incydentu, zazwyczaj stanowi on ogniwo większej serii. I rzeczywiście nie była to pierwsza napaść ze strony tego mężczyzny. Po dokonaniu analizy krwi znalezionej w samochodzie policja stwierdziła, że DNA pokrywa się z próbką sprawcy gwałtu na innym parkingu wielopoziomowym, do którego doszło dwa lata wcześniej w Nottingham, około stu kilometrów na południe od Leeds. Po apelu opublikowanym w programie kryminalnym BBC Crimewatch policja

zdołała powiązać tę sprawę z trzema incydentami sprzed dziesięciu lat – w Bradford, Leeds i Leicester[2].

Mimo to odnalezienie tego seryjnego gwałciciela nie było wcale proste. Obszar wyznaczony przez linie łączące miejsca popełnienia przestępstw liczył 7046 kilometrów kwadratowych – to ogromny szmat lądu. Policja stanęła wobec zatrważającej liczby potencjalnych sprawców: 33 628 mężczyzn, z których każdy będzie musiał zostać albo wykreślony z listy, albo objęty dochodzeniem[3].

Trzeba było przeprowadzić poszukiwania zakrojone na gigantyczną skalę, i to nie pierwszy raz. Dekadę wcześniej policja prowadziła już podobną obławę na gwałciciela, lecz mimo przeszukania 14 153 mieszkań i pobrania ogromnej liczby próbek krwi, włosów i całej masy innego materiału dowodowego dochodzenie ostatecznie utknęło w martwym punkcie. Pewnie nowe śledztwo zakończyłoby się podobnie, gdyby nie poproszono w końcu o pomoc byłego kanadyjskiego gliny Kima Rossma, uzbrojonego w najnowszy algorytm[4].

Rossmo postawił śmiałą tezę. Jego algorytm praktycznie w całości odrzucał ogromny materiał dowodowy zebrany w trakcie dochodzenia, gdyż interesował go wyłącznie jeden czynnik – geografia.

Być może, argumentował Rossmo, nieprzypadkowy jest również wybór m i e j s c a kolejnej napaści; być może to wynik przemyślanej decyzji? Mimo że gwałciciel uderzał na północy i na południu Anglii, Rossmo przypuszczał, że w geografii popełnionych przestępstw kryje się jakaś niezamierzona prawidłowość – prawidłowość na tyle oczywista, że da się ją wykorzystać do zlokalizowania sprawcy. Jego zdaniem istniało spore prawdopodobieństwo, że położenie miast, w których doszło do gwałtów, zdradzi miejsce zamieszkania gwałciciela. Sprawa wampira z Anglii dawała mu szansę na sprawdzenie tej teorii w praktyce.

Operacja Ryś i zraszacz ogrodowy

To jednak nie Rossmo pierwszy wysunął przypuszczenie, że przestępcy mogą nieświadomie pozostawiać geograficzne ślady. Ta idea ma już całkiem długą tradycję, jej początki sięgają lat dwudziestych XIX wieku,

kiedy to prawnik André-Michel Guerry został statystykiem we francuskim Ministerstwie Sprawiedliwości i zaczął gromadzić wszelkie dane dotyczące gwałtów, morderstw i grabieży, których dopuszczono się w różnych rejonach Francji[5].

Dziś zbieranie tego rodzaju informacji to zwykła rutyna, w tamtych czasach jednak matematykę i statystykę stosowano wyłącznie w naukach ścisłych, by za pomocą eleganckich równań opisywać rządzące przyrodą prawa fizyki: ruchy planet w kosmosie czy siły działające w silniku parowym. Nikt wcześniej nie zawracał sobie głowy zbieraniem informacji o zbrodniarzach, bo i nikt nie miał pojęcia, co może być istotne, co można z tego wyliczyć i jak często to robić. Tak więc współcześni pana Guerry'ego mogli się tylko pukać w czoło – na co mu to wszystko? Człowiek to istota silna i z natury niezależna, która chodzi, gdzie chce, i robi, co chce, została bowiem obdarzona wolną wolą; jej zachowanie nie da się w żadnej mierze zamknąć w ramy żadnej statystyki[6].

Sporządzona przez Guerry'ego analiza krajowego spisu przestępców mówiła jednak coś zupełnie innego. Ministerialny statystyk odkrył, że w każdym bez wyjątku regionie Francji da się zauważyć prawidłowości dotyczące tego, kto co robi i jak. Młodzi ludzie popełniali więcej przestępstw niż starsi, mężczyźni więcej niż kobiety, a biedni więcej niż bogaci. Nadzwyczaj intrygujące było też to, że wkrótce stało się jasne, iż prawidłowości te są niezmienne w czasie. Każdy region miał własną statystykę przestępstw, która z przerażającą dokładnością powtarzała się z roku na rok w niemal identycznej liczbie grabieży, gwałtów i mordów. Przewidywalne były nawet metody pozbawiania życia. Guerry i jego współpracownicy mogli zatem wybrać dowolny obszar Francji i z góry określić, jakiej liczby morderstw z użyciem noża, miecza i kamienia lub wskutek uduszenia czy utopienia można się spodziewać w kolejnym roku[7].

Nie można zatem wykluczyć, że coś takiego, jak wolna wola przestępcy, po prostu nie istnieje. Zbrodnie nie są dziełem przypadku – ludzie są przewidywalni. I to właśnie tę przewidywalność – prawie dwa stulecia po odkryciu Guerry'ego – postanowił wykorzystać Kim Rossmo.

Praca Guerry'ego koncentrowała się na prawidłowościach na poziomie kraju i regionu, lecz okazuje się, że nawet na poziomie indywidualnym przestępcy tworzą powtarzalne wzorce geograficzne. Ludzie ci, tak jak my wszyscy, zazwyczaj trzymają się znanych sobie obszarów. Działają lokalnie, a to oznacza, że nawet do tych najcięższych zbrodni dochodzi niedaleko miejsca zamieszkania sprawcy. Im bardziej się oddalasz od miejsca przestępstwa, tym mniejszą masz szansę znalezienia miejsca zamieszkania przestępcy[8] – efekt ten kryminolodzy nazywają czasem „oporem przestrzeni".

Z drugiej strony seryjni przestępcy z reguły nie upatrują sobie ofiar z b y t b l i s k o miejsca zamieszkania, żeby uniknąć nasilonej obecności policji w pobliżu własnego domu czy rozpoznania przez sąsiadów. Tak powstaje „strefa buforowa", która otacza dom przestępcy – to rejon, w którym prawdopodobieństwo popełnienia przez niego przestępstwa jest bardzo małe[9].

Serce algorytmu Rossma stanowiły te dwa kluczowe czynniki – opór przestrzeni i strefa buforowa, zapisane w geografii większości ciężkich przestępstw. Rossmo postanowił matematycznie je zrównoważyć i dzięki temu ustalić potencjalny obszar zamieszkania sprawcy.

Metoda nie jest szczególnie pomocna, jeśli popełniono tylko jedno przestępstwo. Bez wystarczającej liczby danych tak zwany a l g o r y t m p r o f i l o w a n i a g e o g r a f i c z n e g o nie powie ci wiele więcej niż staromodny zdrowy rozsądek. Lecz im więcej będzie przestępstw, tym ostrzejszy staje się obraz; powoli odsłania się mapa miasta z zaznaczonymi obszarami, na których z największym prawdopodobieństwem można znaleźć sprawcę.

Możesz sobie wyobrazić seryjnego przestępcę jako obrotowy zraszacz ogrodowy. Trudno przewidzieć, gdzie upadną pierwsze krople wody, i tak samo nie da się z góry określić, gdzie przestępca zaatakuje za drugim razem. Ale po kilku obrotach zraszacza, kiedy na ziemię upadło już całkiem sporo kropli, stosunkowo łatwo zaobserwować tworzony przez nie wzór i z dużym prawdopodobieństwem oznaczyć lokalizację urządzenia.

Tak właśnie działał algorytm Rossma użyty w operacji Ryś – obławie na seryjnego gwałciciela z Anglii. Zespół śledczy dysponował pięcioma

lokalizacjami przestępstw oraz kilkoma adresami sklepów, w których napastnik kupował alkohol, papierosy i gry wideo na skradzione ofiarom karty kredytowe. Na podstawie wyłącznie tych informacji algorytm oznaczył dwa główne obszary, w których najprawdopodobniej mieszka gwałciciel: Millgarth i Killinbeck, oba na przedmieściach Leeds[10].

Policja miała jeszcze jeden kluczowy dowód w tej sprawie: częściowy odcisk palca pozostawiony przez napastnika na miejscu jednego z wcześniejszych ataków. Próbka była zbyt mała, żeby automatyczny system identyfikacji daktyloskopijnej mógł wyłowić sprawcę z bazy danych skazanych przestępców, tak więc wszelkie badania porównawcze musiał wykonywać uzbrojony w lupę ekspert. Na tym etapie śledztwo trwało już prawie trzy lata i mimo ogromnych wysiłków 180 policjantów z pięciu różnych wydziałów pomysły na przełom zaczynały się wyczerpywać. Każda wskazówka prowadziła do kolejnej ślepej uliczki.

Śledczy postanowili ręcznie sprawdzić wszystkie odciski palców pobrane w dwóch wskazanych przez algorytm obszarach. Najpierw sprawdzono Millgarth, lecz poszukiwania w bazie odcisków tamtejszej komendy policji nie przyniosły rezultatu. Sprawdzono więc Killinbeck – i po 940 godzinach przesiewania tamtejszych kartotek policjanci w końcu ustalili personalia sprawcy, okazał się nim niejaki Clive Barwell.

Barwell miał 42 lata, był żonaty, miał czwórkę dzieci i w przerwie między atakami siedział w więzieniu za napad z bronią w ręku. Obecnie pracował jako kierowca ciężarówki i regularnie jeździł w długie trasy na północ i południe kraju; mieszkał jednak w Killinbeck i często odwiedzał matkę w Millgarth, drugim obszarze wytypowanym przez algorytm[11]. Sam częściowy odcisk palca nie wystarczyłby, żeby zidentyfikować go ponad wszelką wątpliwość, ale przeprowadzone później badanie DNA potwierdziło, że to on popełnił potworne zbrodnie, o które go oskarżano. Policja znalazła wampira. Na rozprawie sądowej w październiku 1999 roku Barwell przyznał się do winy. Sędzia skazał go na osiem kar dożywotniego więzienia[12].

Po zamknięciu sprawy Rossmo miał okazję ocenić, jak dobrze się spisał jego algorytm. Nie podał personaliów przestępcy, wskazał jednak na mapie obszary, na których policja powinna się szczególnie skupić. Gdyby użyto

algorytmu Rossma do priorytetyzacji listy podejrzanych na podstawie ich miejsca zamieszkania – sprawdzając po kolei odciski palców i pobierając od nich próbki DNA – nie byłoby konieczności niepokojenia aż tylu niewinnych osób; policja znalazłaby Clive'a Barwella po sprawdzeniu zaledwie trzy procent obszaru[13].

Nie ulega wątpliwości, że algorytm się sprawdził. Ma też inne plusy: ponieważ priorytetyzuje wyłącznie już istniejącą listę podejrzanych, nie jest obarczony grzechem uprzedzenia opisanym w rozdziale „Wymiar sprawiedliwości". Poza tym nie przekreśla dobrej detektywistycznej roboty, lecz jedynie zwiększa skuteczność dochodzenia – nie zachodzi więc obawa, że policjanci mogliby zacząć wierzyć w niego za bardzo.

Algorytm Rossma jest także niewiarygodnie elastyczny. Od czasów operacji Ryś korzystało z niego ponad 350 wydziałów dochodzeniowych na całym świecie, w tym amerykańskie Federalne Biuro Śledcze i Kanadyjska Królewska Policja Konna. Co więcej, jego możliwości wykraczają daleko poza wykrywanie przestępców: użyty został do wskazania zbiorników wody, gdzie wylęgają się komary (na podstawie lokalizacji przypadków malarii w Egipcie[14]). Jeden z moich obecnych doktorantów na University College London próbuje wykorzystać ten sam algorytm do namierzania fabryk fugasów, wychodząc od lokalizacji miejsc, w których są podkładane. A grupa londyńskich matematyków próbowała nawet wykorzystać go do namierzenia Banksy'ego[15], unikającego rozgłosu artysty ulicznego, na podstawie położenia miejsc, gdzie odnaleziono jego murale.

Przestępstwa, przy których profilowanie geograficzne sprawdza się najlepiej – seryjne gwałty, morderstwa i brutalne napaści – są na szczęście rzadkie. Ogromna większość powszednich naruszeń prawa nie wymaga organizowania takich obław jak ta na Clive'a Barwella. Gdyby algorytmy miały coś zmienić w walce z tą bardziej powszednią przestępczością, musiałyby się posiłkować innym wzorcem geograficznym. Takim, który można by zastosować do miasta jako całości i który potrafiłby wykryć prawidłowości oraz rytmy ulicy czy nawet jej konkretnego rogu – czymś, co każdy dzielnicowy wyczuwa instynktownie. Na szczęście Jack Maple miał właśnie coś takiego.

Mapy przyszłości

Wielu ludzi bało się w latach osiemdziesiątych jeździć nowojorskim metrem. Nie było tam przyjemnie. Wszystko zapaćkane sprejem, wagony cuchnące moczem, perony zawłaszczone przez handlarzy narkotyków, złodziei i bandziorów. Pod ziemią mordowano rocznie około 20 procent niewinnych ludzi, przez co nowojorskie metro było faktycznie jednym z najniebezpieczniejszych miejsc na świecie.

W takich właśnie czasach przyszło pracować w policji komunikacyjnej Jackowi Maple'owi. Maple dochrapał się niedawno awansu na porucznika i przez te wszystkie lata spędzone w metrze miał już serdecznie dosyć uganiania się za przestępcami – wolałby raczej robić wszystko, żeby ograniczyć ich liczbę. Na gruncie tej zawodowej frustracji wykiełkował genialny pomysł.

„Na 17-metrowej ścianie sporządziłem mapę, na której zaznaczyłem każdą stację metra w Nowym Jorku i każdy pociąg – powiedział Maple w wywiadzie w 1999 roku. – Potem kredkami naniosłem na nią każde przestępstwo z użyciem przemocy, każdą napaść i każdą znaczną grabież mienia. Naniosłem sprawy rozwiązane i nierozwiązane"[16].

Nie brzmi to pewnie jak odkrycie koła, niemniej mapy Maple'a, rysowane kredkami na szarym papierze i nazywane później „mapami przyszłości", wywołały w tamtych czasach prawdziwą rewolucję. Nikt przed nim nie wpadł na to, żeby w taki właśnie sposób patrzeć na przestępczość. Kiedy Maple odsunął się od ściany i jednym spojrzeniem ogarnął całą kryminalną nędzę Nowego Jorku, uświadomił sobie, że patrzy na miasto z całkiem nowej perspektywy.

„Zadałem sobie pytanie: dlaczego? Jakie są przyczyny tego, że w danym miejscu dochodzi do akurat takich, a nie innych przestępstw?" Problemem w owych czasach było to, że każde zgłoszenie z metra traktowano jako odizolowany incydent. Jeżeli ktoś zadzwonił z informacją, że tu i tu rozstawiła się grupa agresywnych dilerów narkotyków, którzy natychmiast zapadali się pod ziemię, kiedy na horyzoncie pojawiali się policjanci, w kartotekach nie zachowywały się żadne ślady pozwalające powiązać to zgłoszenie z innymi, które otrzymała policja po powrocie gangu do swojego rewiru. Tymczasem mapy Maple'a pozwalały dokładnie

zlokalizować miejsca ustawicznych problemów z przestępczością, co umożliwiało poszukiwanie przyczyn. „Czy jest tam gdzieś centrum handlowe? Czy to nie dlatego mamy tam tak dużo kradzieży kieszonkowych i rozbojów? Czy tam jest szkoła? Czy to nie dlatego mamy tam zawsze problemy o trzeciej po południu? Czy jest tam gdzieś w pobliżu opuszczony dom? Czy to nie dlatego mamy tam na rogu problemy z dilerami kraku?”[17].

Szukanie odpowiedzi na te pytania stanowiło pierwszy krok w stronę rozwiązania problemów miasta. Tak więc w 1990 roku, kiedy szefem nowojorskiej policji komunikacyjnej został znany z szerokich horyzontów Bill Bratton, Maple pokazał mu swoje „mapy przyszłości” – później korzystali z nich wspólnie, próbując zmienić metro w miejsce bezpieczniejsze dla wszystkich[18].

Bratton też nie był w ciemię bity. Wiedział, że żebranie, sikanie i przeskakiwanie przez bramki obrotowe to poważny problem w metrze, postanowił więc skoncentrować prace policji na przeciwdziałaniu właśnie tym drobnym wykroczeniom, a nie na zajmowaniu się znacznie poważniejszymi przestępstwami – rozbojami i morderstwami, które również przybrały pod ziemią rozmiary epidemii.

Działanie to miało dwa logiczne wytłumaczenia. Po pierwsze, twarde podejście do wszelkich zachowań antyspołecznych w miejscach patogennych stanowiło wyraźny sygnał, że policja nie będzie tolerować przestępczości w żadnej formie – a to mogło doprowadzić do zmiany postrzegania tego, co „normalne”. Po drugie, można z dużym prawdopodobieństwem założyć, że unikanie opłaty za przejazd silnie koreluje z gotowością do popełnienia poważniejszego przestępstwa w metrze, a zatem zatrzymywanie potencjalnych przestępców za przeskakiwanie bramki to skuteczny sposób zapobiegania przestępczości. „Kiedy zaczęliśmy walczyć z gapowiczami, okazało się, że zatrzymujemy na bramkach wielu groźnych, uzbrojonych kryminalistów, którzy mogliby siać pod ziemią prawdziwe spustoszenie”, powiedział Bratton gazecie „Newsday” w 1991 roku[19].

Ta strategia przyniosła efekty. Policja stała się cwańsza, a metro bezpieczniejsze. W latach 1990–1992 mapy Maple’a w połączeniu z taktyką

Brattona pozwoliły na ograniczenie liczby zbrodni w metrze o 27 procent i rozbojów o jedną trzecią[20].

Kiedy Bratton został komisarzem całej nowojorskiej policji, postanowił wprowadzić „mapy przyszłości" Maple'a także na powierzchni. Rozbudowane i poprawione, przeistoczyły się w CompStat, narzędzie śledzenia danych, wykorzystywane obecnie przez wiele wydziałów policji zarówno w Stanach Zjednoczonych, jak i za granicą. Jego podstawą jest nadal prosta zasada Jacka Maple'a: notować, gdzie dochodzi do przestępstw, żeby określić najgorsze siedliska zbrodni w mieście.

Siedliska takie mają z reguły ściśle określone granice. Długofalowe badanie prowadzone w Bostonie przez 28 lat wykazało na przykład, że 66 procent wszystkich rozbojów ulicznych ogranicza się do zaledwie 8 procent ulic miasta[21]. Inne badanie pozwoliło na lokalizację 300 tysięcy miejsc, z których policja w Minneapolis odebrała zgłoszenie o przestępstwie: połowa z nich pochodziła z zaledwie 3 procent obszaru miasta[22].

Ta lokalizacja nie jest jednak niezmienna w czasie – siedliska bezustannie migrują, przesuwając się subtelnie i przekształcając niczym krople oleju na wodzie – niekiedy nawet z dnia na dzień. A gdy w 2002 roku Bratton przeniósł się do Los Angeles, zaczął się zastanawiać, czy nie istnieją jeszcze inne prawidłowości, które mogłyby zdradzać nie tylko miejsce, lecz także czas zbrodni. Czy da się w jakiś sposób spojrzeć dalej, poza przestępstwa, które już popełniono? Zamiast tylko gonić przestępców – co tak frustrowało Maple'a – i walczyć z tym, co już było, umieć także zbrodnie przewidywać?

Sygnał i wzmocnienie

Jeżeli chcemy przewidywać przestępstwa, najlepiej będzie, jak zaczniemy od włamań. Do włamania dochodzi w ściśle określonym miejscu, można więc również w miarę dokładnie określić, kiedy do niego doszło – gdy ofiara była w pracy albo wieczorem, kiedy wyszła do kina. Inaczej rzecz się ma z kradzieżą telefonu na mieście; jest całkiem możliwe, że ofiara zauważy brak komórki dopiero po powrocie do domu. Poza tym większość

włamań zgłasza się policji, dysponujemy więc naprawdę dobrą, bogatą bazą danych; znacznie trudniej zbudować podobną bazę dla przestępstw związanych, na przykład, z handlem narkotykami.

Poza tym włamywacze mają jedną cechę wspólną z seryjnymi mordercami i gwałcicielami, których badał Rossmo: preferują miejsca, które dobrze znają. Wiemy już, że twoje mieszkanie jest dużo bardziej narażone na włamanie, jeżeli mieszkasz przy ulicy, którą włamywacz regularnie chodzi – na przykład do pracy albo do szkoły[23]. Wiemy także[21*], że włamywacze mają swoje upodobania, jeśli chodzi o ruchliwość okolicy: unikają zazwyczaj ulic uczęszczanych, a jednocześnie lubią miejsca odwiedzane przez beztroskich nietutejszych, ponieważ w takich warunkach łatwiej wtopić się w tłum (chyba że na ulicy jest sporo hałaśliwych miejscowych mogących pełnić nieświadomie funkcję strażników)[24].

Ale to dopiero pierwszy z dwóch komponentów atrakcyjności twojego domu dla włamywacza. To prawda, są czynniki niezmienne w czasie, takie jak miejsce, w którym mieszkasz, czy ruchliwość twojej ulicy – czynniki te bezustannie „sygnalizują" atrakcyjność twojej nieruchomości, czyniąc z niej potencjalny cel. Zanim jednak sprzedasz swój dom i przeniesiesz się do cichego zaułka z monitoringiem, musisz mieć świadomość, że skupiska przestępczości nie są stałe. Tym bardziej że drugi komponent atrakcyjności twojego domu jest raczej na pewno ważniejszy od pierwszego, a jego wartość zależy od tego, co dokładnie dzieje się właśnie teraz w pobliżu twojego domu. To zjawisko nazywamy „wzmocnieniem".

Jeżeli doświadczyłeś kiedyś dwóch włamań w krótkim odstępie czasu, aż za dobrze wiesz, na czym polega efekt wzmocnienia. Policjanci po pierwszym włamaniu powiedzieli ci na pewno, że przestępcy mają skłonność do powracania do tego samego miejsca – co oznacza, że bez względu na to, gdzie mieszkasz, najbardziej narażony na powtórkę z rozrywki jesteś w ciągu kilku dni od pierwszego włamania; w tym czasie ryzyko drugiego włamania może wzrosnąć nawet dwunastokrotnie[25].

Istnieje kilka powodów, dla których włamywacze mogą się zdecydować na powrót do twojego domu. Mogli już poznać jego rozkład i wiedzą, gdzie trzymasz kosztowności (zresztą takie rzeczy jak telewizor i komputery też

są bardzo szybko zastępowane nowymi), albo rozpracowali zamki w drzwiach i drogi szybkiej ewakuacji; a może wpadł im w oko jakiś przedmiot, którego nie zdołali wynieść za pierwszym razem. Tak czy owak, efekt wzmocnienia nie odnosi się wyłącznie do ciebie. Badacze odkryli, że ryzyko włamania do sąsiadów tuż po okradzeniu twojego mieszkania również znacząco wzrasta, podobnie jak ryzyko włamania do ich sąsiadów, sąsiadów ich sąsiadów i tak dalej, aż do końca ulicy.

Wzmocnienia można postrzegać jak fajerwerki rozbłyskujące w różnych częściach miasta. Im dalej od pierwszej iskry, tym efekt wzmocnienia staje się słabszy i z czasem zanika, by po dwóch miesiącach – chyba że nowe przestępstwo ponownie „wzmocni sygnał" tej samej ulicy – zgasnąć zupełnie[26].

Sygnał i wzmocnienie mają nawet kuzynów w świecie przyrody. Są to trzęsienia ziemi. To prawda, nie da się dokładnie przewidzieć, gdzie i kiedy nastąpi pierwszy wstrząs (choć wiemy, że niektóre miejsca są bardziej na nie podatne od innych). Ale gdy tylko ziemia zacznie gdzieś drżeć, da się przewidzieć, gdzie i ilu wstrząsów wtórnych można się spodziewać – przy czym ryzyko ich wystąpienia jest największe w miejscu pierwszego trzęsienia i zmniejsza się z oddalaniem od niego, by w końcu zaniknąć zupełnie.

Prawidłowości trzęsień ziemi i włamań po raz pierwszy powiązano ze sobą w tym czasie, kiedy Bratton kierował policją w Los Angeles. A ponieważ policja w Los Angeles dałaby bardzo wiele za możliwość przewidywania zbrodni, podpisała umowę z grupą matematyków z Uniwersytetu Kalifornijskiego i przekazała im wszystkie dane, jakie tylko miała w swojej dyspozycji: 13 milionów przestępstw popełnionych w ciągu 80 lat. Choć w owym czasie kryminolodzy wiedzieli już o zjawiskach sygnału i wzmocnienia, dzięki analizie prawidłowości występujących w udostępnionych im danych kalifornijscy matematycy jako pierwsi odkryli, że równania matematyczne, które tak pięknie przewidują ryzyko wstrząsów sejsmicznych i wstrząsów wtórnych, można również stosować do przewidywania przestępstw i „przestępstw wtórnych" – i to nie tylko włamań. Tak powstało narzędzie do przewidywania wszystkiego: od kradzieży samochodów, przez przestępstwa z użyciem przemocy, po zwykłe akty wandalizmu.

To było prawdziwe trzęsienie ziemi w policji. Zamiast ogólnikowo twierdzić, że okolica miasta, gdzie niedawno doszło do przestępstwa, jest „bardziej narażona" na powtórzenie się scenariusza, teraz można było to ryzyko wyliczyć dla każdej ulicy z osobna. A mając statystyczną pewność, że dana okolica miasta będzie w daną noc celem włamywaczy, łatwo już było napisać algorytm mówiący policji, na które ulice trzeba zwrócić szczególną uwagę.

Tak narodził się PredPol (od ang. *PREDictive POLicing*, Prewencja Prognostyczna).

Informator

Niewykluczone, że wiesz już o istnieniu PredPola. Od inauguracji w 2011 roku był tematem tysięcy doniesień prasowych – zazwyczaj pod nagłówkiem mającym jakiś związek z filmem *Raport mniejszości* z Tomem Cruise'em. PredPol to dziś Kim Kardashian świata algorytmów: bije rekordy popularności, wzbudza w mediach tsunami krytyki, lecz nikt tak naprawdę nie wie, czym się zajmuje.

Zanim jednak oczami wyobraźni zobaczysz leżących w basenie wróżów wykrzykujących swoje ostrzeżenia, powściągnę nieco twoją fantazję. PredPol nie typuje potencjalnych sprawców przed popełnieniem zbrodni. Nie może w ogóle wskazywać nikogo – interesuje go tylko i wyłącznie geografia przestępstwa. To prawda, używałam tutaj dość często słowa „przewidywać", lecz algorytm nie potrafi przepowiadać przyszłości. To nie jest kryształowa kula. PredPol może jedynie przewidywać ryzyko wystąpienia określonych zdarzeń, a nie same zdarzenia – a to subtelna, lecz ważna różnica.

Wyobraź sobie, że algorytm to bukmacher: wielka grupa policjantów tłoczy się nad mapą miasta i obstawia, gdzie tej nocy ktoś popełni przestępstwo, a PredPol oblicza kursy. To taki anonimowy informator, który daje policji cynk, że te i te ulice będą dzisiaj „gorące" – i oznacza je na mapie czerwonym kwadracikiem.

Teraz trzeba sobie zadać pytanie, czy opłaci się obstawić miejsca wytypowane przez informatora. Żeby sprawdzić skuteczność algorytmu[27], kazano mu się zmierzyć z najlepszymi analitykami

kryminalistyki w dwóch odrębnych eksperymentach; jeden przeprowadzono w hrabstwie Kent w południowej Anglii, a drugi w południowo-zachodniej strefie Los Angeles. Były to dwa pojedynki jeden na jednego. Algorytm i ekspert mieli identyczne zadanie: ustawić na mapie 20 kwadracików reprezentujących obszar 150 metrów kwadratowych w terenie – tam gdzie ich zdaniem w ciągu kolejnych 12 godzin zostanie popełnionych najwięcej przestępstw.

Zanim przejdziemy do wyników, musimy zdać sobie sprawę, jak piekielnie trudne było to zadanie. Gdyby tobie czy mnie kazano zrobić to samo – przy założeniu, że nie dysponujemy rozległą wiedzą o kryminalnym świecie hrabstwa Kent i Kalifornii – nasze rezultaty pewnie nie byłyby wcale lepsze od tych uzyskanych po losowym rozrzuceniu kwadracików na mapie. Dodam, że 20 użytych w eksperymencie kwadracików tworzy łącznie beznadziejnie mały obszar – jedną tysięczną część hrabstwa Kent[28] – a co 12 godzin trzeba by zaczynać typowanie od początku. Przy tego rodzaju przypadkowym rozrzucie moglibyśmy oczekiwać trafnej „przepowiedni" rzadziej niż raz na sto przestępstw[29].

Eksperci wypadli znacznie lepiej. W Los Angeles analityk zdołał trafnie przewidzieć miejsca wystąpienia 2,1 procent przestępstw[30], a eksperci brytyjscy poradzili sobie jeszcze lepiej, uzyskując średnią 5,4 procent[31] – wynik imponujący, tym bardziej że ich mapa obejmowała obszar mniej więcej dziesięciokrotnie większy niż mapa Los Angeles.

Ale wszystko to blednie wobec wyniku algorytmu, który w Los Angeles trafnie przewidział ponad dwa razy więcej przestępstw niż człowiek, a podczas jednego z etapów brytyjskich niemal jedno na pięć przestępstw popełniono w granicach czerwonych kwadracików wytypowanych dzięki matematyce[32]. PredPol nie jest kryształową kulą, ale jako pierwszy w historii przewidział zbrodnię z taką dokładnością.

Przepowiednie a praktyka

Jest jednak pewien problem. Algorytm jest naprawdę dobry w przepowiadaniu, gdzie zostanie popełnione przestępstwo w ciągu kolejnych 12 godzin, policja ma jednak trochę inny cel – o g r a n i c z y ć

liczbę przestępstw popełnionych w ciągu kolejnych 12 godzin. Pozostaje zatem pytanie: co dalej robić z przepowiednią algorytmu?

Mamy przynajmniej kilka możliwości. W przypadku włamania można ustawić na miejscu zdarzenia kamery monitoringu albo policjantów w cywilnym ubraniu i złapać złodziei na gorącym uczynku. Byłoby jednak lepiej dla wszystkich, gdyby nasze wysiłki zmierzały do zapobiegania przestępstwom. Bo w końcu co lepsze: paść ofiarą przestępstwa, którego sprawca został złapany, czy w ogóle do niego nie dopuścić?

Można ostrzec mieszkańców wytypowanej ulicy, że ich własność jest zagrożona, można zaproponować im wymianę zamków na lepsze, zainstalowanie alarmu antywłamaniowego albo programatorów przy światłach w mieszkaniu, żeby z zewnątrz wyglądało na to, że ktoś jest w środku. Zrobiono tak podczas pewnego eksperymentu w Manchesterze w 2012 roku[33] i zdołano ograniczyć liczbę włamań o ponad jedną czwartą. Jest jednak mały minus: badacze obliczyli, że podobna strategia „utwardzania targetu" będzie nas kosztować około 3925 funtów na jedno włamanie, któremu zdołamy zapobiec[34]. Spróbuj teraz sprzedać to rozwiązanie policji w Los Angeles, która ma co roku ponad 15 tysięcy włamań[35].

Inna opcja, która niewiele odbiega od tradycyjnych metod działania policji, to strategia „prewencji punktowej" (ang. *cops on dots*, dosł. „gliny w punktach").

„W dawnych czasach – powiedział mi Steyve Colgan, emerytowany oficer londyńskiej Policji Metropolitalnej – [patrole] były tylko geograficzne, dostawałeś mapę, rozbijałeś ją na kwadraty i dzieliłeś między ludzi. Ty masz ten kwadrat, a ty masz ten kwadrat. I już". Kłopot w tym, że – jak obliczono w pewnym brytyjskim badaniu – pieszy policjant patrolujący losowo wyznaczony kwadrat ma szansę znaleźć się w odległości stu metrów od właśnie okradanego mieszkania zaledwie raz na osiem lat[36].

Natomiast w strategii prewencji punktowej po prostu wysyłasz patrol w miejsca wytypowane przez algorytm. Oczywiście po to, żeby policjanci rzucali się w oczy we właściwym czasie i miejscu, jest bowiem bardzo prawdopodobne, że skłoni to przestępców do odstąpienia od swoich

zamiarów, a w najgorszym razie – pozwoli policji błyskawicznie zareagować.

Właśnie tak się stało w hrabstwie Kent. Podczas drugiej fazy eksperymentu, na początku nocnej zmiany, sierżant wydrukował mapy z czerwonymi kwadracikami oznaczającymi miejsca szczególnie niebezpieczne tej nocy. W każdej spokojniejszej chwili policjanci mieli podjechać do najbliższego punktu zapalnego, wysiąść z samochodu i zrobić mały obchód terenu.

Tamtego wieczoru w okolicy, do której policjanci nigdy się nie zapuszczali, wpadli na ulicy na kobietę z dzieckiem. Pochodziła ze wschodniej Europy, żyła w patologicznym związku i właśnie kilka minut wcześniej jej dziecko było molestowane. Sierżant, który dyżurował tej nocy w komendzie, potwierdził: „Znaleźliśmy tych ludzi, bo byli w kwadraciku PredPola”[37]. Podejrzany został aresztowany w tej samej okolicy przed końcem nocnej zmiany.

Ta matka i jej dziecko nie byli jedynymi osobami, którym pomógł algorytm w okresie testowania prewencji punktowej: przestępczość w Kencie ogólnie zmalała o 4 procent. Podobne badania w Stanach Zjednoczonych (przeprowadzone przez samą firmę PredPol) podają jeszcze wyższe liczby. W dzielnicy Foothill w Los Angeles przestępczość spadła o 13 procent w ciągu pierwszych czterech miesięcy używania algorytmu, gdy tymczasem w pozostałych dzielnicach miasta, gdzie policja stosowała bardziej tradycyjne metody, ogólna przestępczość wzrosła o 0,4 procent. A raport z Alhambry, miasta w Kalifornii niedaleko Los Angeles, mówił o niewiarygodnym 32-procentowym zmniejszeniu się liczby włamań i 20-procentowym zmniejszeniu się liczby kradzieży pojazdów po wprowadzeniu algorytmu w styczniu 2013 roku[38].

Te liczby robią wrażenie, trudno jednak stwierdzić, czy całą zasługę można przypisać PredPolowi. Jak powiedział mi Toby Davies, matematyk i ekspert kryminalistyki z Uniwersytetu Kalifornijskiego: „Możliwe, że już sam fakt, iż policjanci jeździli w różne miejsca, wysiadali z radiowozów i chodzili po ulicach, przyczynił się do spadku (liczby popełnianych przestępstw)”.

No i jeszcze jedna kwestia. Jeżeli jest tak, że im intensywniej szukasz działalności kryminalnej, tym bardziej jest prawdopodobne, że się jej dopatrzysz, to wówczas już samo wysyłanie w teren policji może wpływać na statystyki: „Kiedy policjanci są na miejscu – powiedział mi Davies – wykrywają więcej przestępstw, niż gdyby ich tam nie było. Nawet jeśli w dwóch różnych miejscach dochodzi do tej samej liczby przestępstw, policja wykryje ich więcej tam, gdzie miała swoich ludzi, niż tam, gdzie ich nie było".

To oznacza, że istnieje potencjalnie bardzo duży minus stosowania strategii prewencji punktowej. Kiedy wysyłasz policjantów do walki z przestępczością w miejsca wyznaczone przez algorytm, możesz sobie zafundować sprzężenie zwrotne.

Powiedzmy, że pewna biedna dzielnica ma wysoki poziom przestępczości, a algorytm przewiduje, że poziom ten jeszcze wzrośnie. W dzielnicy pojawia się zatem więcej policjantów, którzy oczywiście wykrywają więcej przestępstw, niż wykrywali do tej pory. Algorytm przewiduje więc, że liczba przestępstw w tej dzielnicy będzie dalej rosła, policja wysyła tam jeszcze więcej ludzi – i kółko się zamyka. Tego rodzaju sprzężenia zwrotne mogą sprawiać szczególne kłopoty w prewencji przestępstw i wykroczeń wiązanych najczęściej z biednymi dzielnicami miast: żebractwem, włóczęgostwem i stosowaniem miękkich narkotyków na własne potrzeby.

W Wielkiej Brytanii, gdzie przedstawiciele niektórych grup społecznych regularnie się skarżą na brak policji na ulicach, w skupieniu uwagi sił porządkowych na określonych obszarach nie byłoby pozornie nic złego. Ale nie wszyscy żyją w dobrych stosunkach z policją. „Ludzie widzący policjanta spacerującego pod ich domem dzień w dzień mogą mieć uprawnione poczucie, że są inwigilowani i dyskryminowani, nawet jeśli nikt tam nie robi nic złego, a policjant naprawdę tylko sobie spaceruje – powiedział Davies. – Można by wręcz powiedzieć, że człowiek ma prawo do tego, by nie wywierano na niego bezustannie presji, by nie był pod ciągłą obserwacją policji".

Ja też się skłaniam ku tej interpretacji.

Moim zdaniem dobry algorytm p o w i n i e n być tak wyregulowany, żeby uwzględniał strategie stosowane przez policję. Teoretycznie istnieją

sposoby powstrzymania algorytmu przed nieproporcjonalnie częstym oznaczaniem określonych okolic, na przykład losowe wysyłanie patroli nie tylko do stref wysokiego, lecz także średniego ryzyka. Niestety, nie zdołamy nigdy stwierdzić z całą pewnością, czy PredPol potrafi unikać opisanych wyżej sprzężeń zwrotnych, ponieważ jest to algorytm własnościowy; jego kod nie jest ogólnie dostępny i nikt nie wie dokładnie, jaka jest zasada jego działania.

PredPol to nie jedyny taki program na rynku. Do jego konkurentów należy HunchLab, który działa w oparciu o rozmaite statystyki lokalne, takie jak liczba stwierdzonych przestępstw i zgłoszeń na policję czy spis ludności (ale też znacznie bardziej zastanawiające dane w rodzaju faz Księżyca). Za algorytmem HunchLabu nie kryje się żadna teoria, maszyna nie próbuje ustalać, dlaczego niektóre obszary są bardziej dotknięte przestępczością niż inne; po prostu informuje o prawidłowościach, które znajduje w przedstawianych mu danych i w rezultacie potrafi wiarygodnie przewidywać więcej rodzajów przestępstw niż PredPol (który bazuje na teoriach tworzenia wzorców geograficznych przez groźnych przestępców). Ponieważ jednak HunchLab, podobnie jak PredPol, stanowi chronioną własność intelektualną, jest praktycznie niemożliwe, żeby osoba z zewnątrz mogła sprawdzić, czy algorytm nie wywołuje niezamierzonej dyskryminacji określonych grup ludności[39].

Innym nieprzejrzystym algorytmem przewidującym jest lista strategicznych podejrzanych (ang. *Strategic Subject List*), używany przez policję w Chicago[40]. Ten algorytm ma zupełnie inne podejście do przestępczości. Nie skupia się na geografii, lecz stara się przewidzieć, kto będzie zamieszany w przestępstwo z użyciem broni. Na podstawie rozmaitych czynników tworzy listę specjalną (ang. *heat list*) z nazwiskami osób, które z wielkim prawdopodobieństwem mogą w najbliższej przyszłości stać się zarówno sprawcami, jak i ofiarami strzelaniny. Teoria jest bez zarzutu: dzisiejsze ofiary są często jutrzejszymi katami. Intencje też nie są złe: policjanci odwiedzają ludzi wytypowanych przez komputer, proponują im udział w programach interwencyjnych i pomagają wejść na dobrą drogę.

Istnieją jednak obawy, że obietnice twórców Strategic Subject List nie pokrywają się z rzeczywistością. Niedawne badanie przeprowadzone przez

organizację non profit o nazwie RAND Corporation wykazało, że pojawienie się na liście generowanej przez algorytm nie miało żadnego wpływu na stopień prawdopodobieństwa udziału danej osoby w strzelaninie[41] – oznaczało natomiast zwiększenie prawdopodobieństwa aresztowania. We wnioskach autorzy raportu napisali, iż działo się tak przypuszczalnie dlatego, że kiedy rzeczywiście doszło do jakiejś strzelaniny, policjanci traktowali listę specjalną po prostu jak listę podejrzanych.

Policyjne algorytmy prognostyczne są bez wątpienia obiecujące, a osoby odpowiedzialne za ich powstanie z pewnością działają w dobrej wierze i mają czyste intencje. Obawy o uprzedzenia i dyskryminację są jednak uzasadnione, a dla mnie pytania te są zbyt ważne dla społeczeństwa sprawiedliwego, żebyśmy mogli ot tak, po prostu przyjąć zapewnienia służb, że będą korzystać z komputerów na uczciwych zasadach. To tylko jeden z wielu przykładów na to, jak bardzo potrzeba nam niezależnych ekspertów i organów regulacyjnych – ktoś musi pilnować, żeby plusy stosowania algorytmów rzeczywiście przeważały nad minusami.

Potencjalne szkody są trudne do oszacowania. Jak już się przekonaliśmy na kilku innych przykładach, algorytmy potrafią upozorować niewłaściwy wynik na prawdę ostateczną. A konsekwencje mogą być dramatyczne. Nic nie staje się prawdą tylko dlatego, że tak chce komputer.

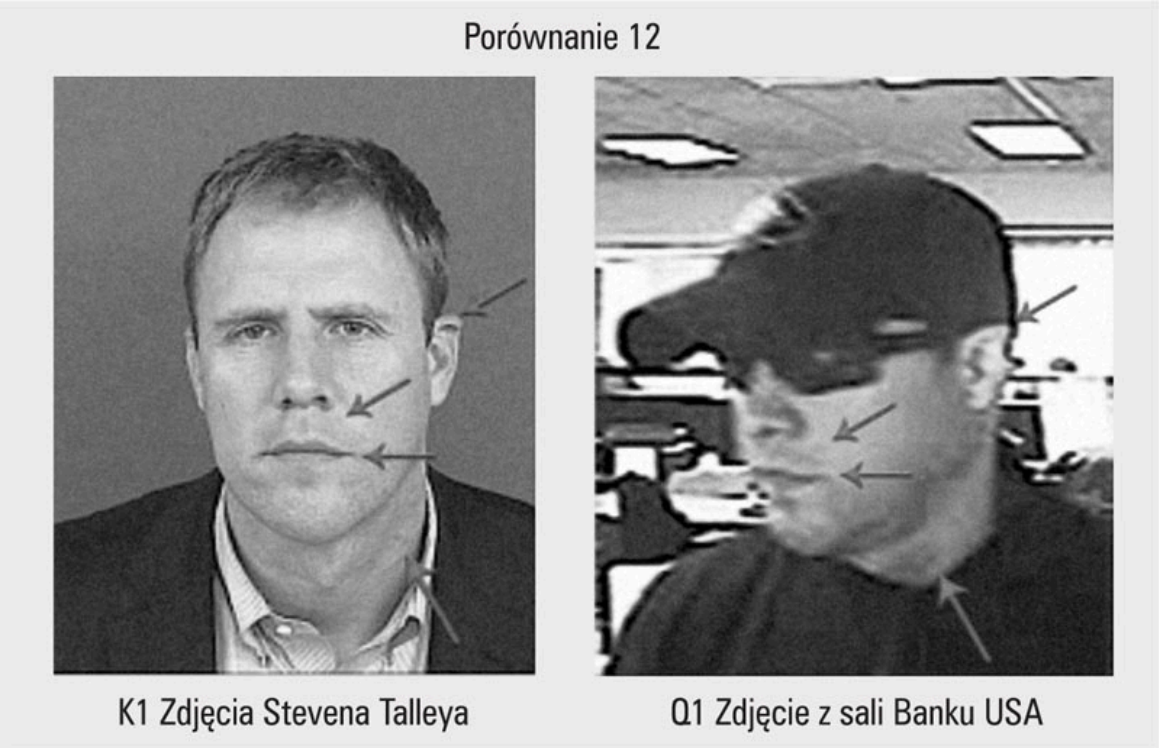
Za kogo ty się masz?

Był rok 2014. Steve Talley spał w swoim domu w południowym Denver, kiedy usłyszał pukanie do drzwi[42]. Kiedy otworzył, zobaczył nieznajomego, który przeprosił i powiedział, że niechcący uderzył w jego samochód, chciał więc, żeby ten wyszedł przed dom i ocenił szkody. Talley się zgodził i kiedy przykucnął, żeby się przyjrzeć drzwiom kierowcy[43], eksplodował granat hukowy. Trzech mężczyzn w czarnych kamizelkach i hełmach przygniotło go do ziemi. Jeden stanął na jego twarzy. Drugi krępował mu ręce, a trzeci zaczął go okładać kolbą karabinu.

Obrażenia Talleyego były poważne. Do wieczora lekarz stwierdził uszkodzenie nerwów, zakrzepy i złamanie prącia[44]. „Nie miałem nawet

pojęcia, że można złamać prącie – powiedział później Talley dziennikarce z «The Intercept». – W pewnym momencie zacząłem wzywać policję. Wtedy do mnie dotarło, że to właśnie gliny mnie biją”[45].

Steve Talley został aresztowany pod zarzutem napadu na dwa okoliczne banki. W drugim jedną z ofiar był policjant i jak przypuszcza Talley, to właśnie dlatego został tak brutalnie potraktowany podczas zatrzymania. „Krzyczałem na nich, że chyba zwariowali – wspomina. – Krzyczałem, że to nie ja!”.



Talley nie kłamał. Za jego aresztowanie odpowiada uderzające podobieństwo do poszukiwanego przestępcy.

Jak się okazało, pierwszy sygnał policja dostała od dozorcy kamienicy, w której Talley mieszkał – po tym jak mężczyzna zobaczył zdjęcia sprawcy w lokalnych mediach – ostateczny werdykt jednak wydał agent FBI, który przeanalizował zapis monitoringu za pomocą programu rozpoznawania twarzy[46] i orzekł, że „poszukiwany osobnik na zdjęciach wygląda jak Talley”[47].

Talley miał żelazne alibi, ale za sprawą ekspertyzy agenta FBI jeszcze rok musiał walczyć o całkowite oczyszczenie swojego imienia. W tym czasie

przez niemal dwa miesiące był trzymany w więzieniu o najostrzejszym rygorze – dopóki na światło dzienne nie wypłynęły dowody, które umożliwiły mu wyjście na wolność. W rezultacie Talley nie mógł pracować i zanim jego gehenna dobiegła końca, stracił pracę, mieszkanie i możliwość widywania dzieci. A wszystko to było bezpośrednim skutkiem owego fałszywego rozpoznania.

Podwójne widzenie

Algorytmy rozpoznawania twarzy stają się coraz powszechniejsze we współczesnej policji. Wystarczy pokazać im zdjęcie, film albo kadr z kamery 3D, a zlokalizują na materiale twarz, pomierzą ją i porównają z istniejącą bazą zdjęć, która pomoże im ustalić tożsamość poszukiwanej osoby.

W Berlinie algorytmy rozpoznawania twarzy zdolne zidentyfikować podejrzanych o terroryzm ćwiczy się na tłumach przetaczających się przez dworce kolejowe[48]. W samym tylko stanie Nowy Jork użycie takich algorytmów doprowadziło od 2010 roku do aresztowania ponad czterech tysięcy osób podejrzanych o oszustwa i kradzieże tożsamości[49]. A na ulicach Wielkiej Brytanii kamery zamontowane na pojazdach przypominających nieco podrasowane samochody Google StreetView bezustannie sprawdzają podobieństwo przechodniów do zdjęć figurujących w bazie osób poszukiwanych[50]. Swój pierwszy sukces te automatyczne psy gończe odnotowały w czerwcu 2017 roku, kiedy jeden z nich wypatrzył na ulicy w południowej Walii mężczyznę poszukiwanego przez policję[51].

Nasze bezpieczeństwo często zależy od zdolności rozpoznawania twarzy. Ale pozostawienie tego zadania w rękach ludzi może być ryzykowne. Weźmy funkcjonariuszy straży granicznej i personel lotniskowy. W niedawnym badaniu, w którym odtworzono warunki panujące podczas kontroli bezpieczeństwa na lotnisku, pracownicy zawodowo zajmujący się rozpoznawaniem twarzy nie zdołali wyłapać osób podróżujących z cudzym dokumentem tożsamości aż w 14 procentach przypadków, natomiast błędnie zakwestionowali 6 procent dokumentów

z autentycznym zdjęciem osoby kontrolowanej[52]. Nie wiem jak was, ale mnie te liczby, delikatnie mówiąc, trochę niepokoją – wystarczy sobie uzmysłowić, ilu pasażerów przewija się każdego dnia przez Heathrow.

Jak się przekonamy, algorytmy rozpoznawania twarzy są w tej pracy zdecydowanie lepsze od ludzi. Kiedy jednak używa się ich do polowania na kryminalistów, gdzie konsekwencje fałszywej identyfikacji są tak poważne, nasuwa się kilka istotnych pytań. Jak łatwo można pomylić ze sobą dwie różne osoby? Jak wielu z nas ma gdzieś na świecie, tak jak Steve Talley, sobowtóra z bogatą przeszłością kryminalną?

Autorka badania przeprowadzonego w 2015 roku stwierdziła, że szanse na to, byśmy mieli gdzieś w rzeczywistym świecie sobowtóra (niekoniecznie notowanego przez policję), są bliskie zeru. Teghan Lucas z University of Adelaide zadała sobie mnóstwo trudu i pomierzyła osiem różnych cech na zdjęciach czterech tysięcy osób, a ponieważ nie zdołała znaleźć ani jednej pary o identycznych miarach, doszła do wniosku, że szanse na to, by dwie osoby miały dokładnie taką samą twarz, są mniejsze niż jeden do biliona[53]. W świetle tych badań trzeba by stwierdzić, że Talley miał naprawdę kosmicznego pecha. A jeśli wziąć pod uwagę również to, że jego zły sobowtór – jeden z biliona! – mieszkał niedaleko o r a z był kryminalistą, powinniśmy być święcie przekonani, że upłyną co najmniej dziesiątki tysięcy lat, zanim kolejny wielki pechowiec będzie musiał przeżyć podobny koszmar.

A jednak mamy powody przypuszczać, że coś jest z tymi liczbami nie tak. Trudno sobie wyobrazić spotkanie z osobą o identycznej twarzy, niemniej anegdoty o niespokrewnionych ze sobą bliźniakach słyszy się znacznie częściej, niżby to wynikało z badań Lucas.

Weźmy przykład Neila Douglasa, który po wejściu na pokład samolotu do Irlandii stwierdził, że na sąsiednim fotelu siedzi jego sobowtór. Selfie, które sobie zrobili obaj panowie – z widocznymi w tle roześmianymi pasażerami – błyskawicznie rozeszło się po internecie i rudzi brodacze z całego świata zaczęli przysyłać swoje zdjęcia, a wszyscy byli do nich mniej lub bardziej podobni. „Mam wrażenie, że w pewnym momencie była nas mała armia", powiedział Neil dziennikarce BBC[54].

Mogłabym nawet dołożyć do tego własną historię. Kiedy miałam 22 lata, koleżanka pokazała mi zdjęcie, które zobaczyła na stronie pewnego zespołu muzycznego na portalu Myspace. Był to w rzeczywistości kolaż zdjęć zrobionych na koncercie, na którym mnie nie było. Na wszystkich widniały twarze świetnie bawiących się osób, z których jedna była niesamowicie podobna do mnie. Aż musiałam się upewnić, że tamtej nocy nie przesadziłam z alkoholem i zupełnie o tym nie wiedząc, nie zawędrowałam na imprezę, z której nie pozostały mi żadne wspomnienia, napisałam więc e-maila do wokalisty kapeli, a ten potwierdził moje przypuszczenia: mam sobowtóra, który uwielbia technopop i ma znacznie bogatsze życie towarzyskie niż ja.

Tak więc Talley, Douglas i ja mamy co najmniej po jednym sobowtórze, a możliwe, że jest ich więcej. To już trzy osoby w populacji liczącej 7,5 miliarda! Nie zaczęliśmy nawet badać sprawy na poważnie, a już bijemy na głowę wyliczony przez Lucas stosunek jeden do biliona.

Istnieje wytłumaczenie tak dużej rozbieżności: wszystko zależy od tego, jak badacz zdefiniuje „identyczność". W swoich badaniach Lucas przyjęła założenie, że wymiary twarzy dwóch osób muszą być sobie równe. I choć Neil i jego sobowtór są niewiarygodnie do siebie podobni, to jeśli tylko ich

nozdrza albo płatki uszne różnią się od siebie zaledwie o milimetr, nie spełniają kryteriów identyczności narzuconych przez Lucas.

Lecz nawet jeżeli będziesz porównywać dwa zdjęcia tej samej osoby, dokładne wymiary nie pokażą ci, jak każdy z nas bezustannie się zmienia – starzejemy się, chorujemy, możemy być zmęczeni, mamy różną mimikę, no i nie zapominajmy o zniekształceniach, jakim podlega obraz w aparacie fotograficznym. Spróbuj ująć zasadnicze rysy twarzy w milimetrach, a przekonasz się, że każdy z nas ma tyle odmian własnej twarzy, ile jest ludzi na Ziemi. Krótko mówiąc: same wymiary to za mało, żeby odróżnić jedną twarz od drugiej.

I choć Neil i jego sobowtór z samolotu nie są doskonale identyczni, potrafię sobie łatwo wyobrazić sytuację, w której mylę ich ze sobą na zdjęciu. Podobnie było w sprawie Talleya – biedny Steve nawet nie był a ż t a k podobny do prawdziwego bandyty, a mimo to eksperci z FBI uznali, że zdjęcia obu mężczyzn przedstawiają jednego i tego samego człowieka, wskutek czego Talleya oskarżono o przestępstwo, którego nie popełnił, i osadzono w celi o najostrzejszym rygorze.

Jak wykazało badanie z funkcjonariuszami straży granicznej, zdumiewająco łatwo jest pomylić ze sobą nieznajome twarze – wystarczy, że wydadzą się podobne na pierwszy rzut oka. Niesamowite, jak kiepsko ludzie sobie radzą z rozpoznawaniem obcych. To właśnie dlatego jedna z moich znajomych stwierdziła, że z trudem wysiedziała do końca pięknego filmu *Dunkierka* Christophera Nolana – bo nie mogła odróżnić od siebie aktorów. Właśnie dlatego nieletni amatorzy alkoholu lubią „pożyczać" dowody tożsamości od starszych kolegów. Założona przez amerykańskich prawników organizacja non profit o nazwie Innocence Project ocenia, że ponad 70 procent niesłusznych oskarżeń bierze się z błędnych wskazań i zeznań naocznych świadków[55].

A jednak… Chociaż naoczny świadek mógłby łatwo pomylić Neila z jego towarzyszem podróży, matka rudego brodacza z pewnością nie miałaby najmniejszych problemów z rozpoznaniem na zdjęciu swojego syna. Jeśli chodzi o ludzi, których znamy, jesteśmy mistrzami świata w rozpoznawaniu twarzy, nawet gdy w grę wchodzą prawdziwe sobowtóry: moglibyśmy nie odróżnić od siebie bliźniaków jednojajowych

znanych nam tylko z widzenia, ale jeżeli to nasi dobrzy znajomi, nie mielibyśmy żadnych problemów.

I tu dochodzimy do kwestii zasadniczej: podobieństwo ukrywa się w oku patrzącego. Bez ściśle określonej definicji, czym podobieństwo jest, nie można zmierzyć, j a k b a r d z o różnią się dwie twarze – nie istnieją punkty wyznaczające granice identyczności. Nie da się zdefiniować, co znaczy być sobowtórem, nie da się powiedzieć, jak „zwyczajna" jest dana twarz, ani też – co szczególnie ważne – nie jesteśmy w stanie określić, do jakiego stopnia jest prawdopodobne, że dowolne dwa zdjęcia zostały zrobione tej samej osobie.

To oznacza, że rozpoznawanie twarzy jako metoda identyfikacji nie może się równać testom DNA. Stosowane w kryminalistyce badania genetyczne koncentrują się na określonych fragmentach genomu – tych, o których wiemy, że bardzo się różnią u poszczególnych ludzi. Kluczem jest właśnie stopień owej rozbieżności: jeżeli sekwencja DNA w próbce tkanki znalezionej na miejscu zbrodni jest identyczna z sekwencją DNA w próbce pobranej od podejrzanego, można obliczyć stopień prawdopodobieństwa, z jakim obie tkanki należą do jednej i tej samej osoby. Oznacza to również, że można dokładnie określić, jak bardzo jest prawdopodobne, że znajdzie się pechowiec, który będzie miał identyczne DNA w badanym zakresie genomu[56]. Im większej liczby markerów się użyje, tym mniejsze będzie ryzyko skazania niewinnego człowieka; wybierając liczbę markerów do testów DNA, wymiar sprawiedliwości każdego państwa na świecie sam ustala dokładność metody[57].

Mimo że nasze twarze wydają się tak nierozerwalnie związane z naszą tożsamością, bez znajomości wariacji gatunkowej człowieka praktyka rozpoznawania przestępców na podstawie twarzy nie będzie się mieścić w ramach ścisłości naukowej. Jeśli chodzi o rozpoznawanie ludzi na podstawie zdjęć, to – cytując fragment prezentacji zespołu kryminalistycznego FBI – „brak statystyki oznacza jedno: ostatecznie wszelkie wnioski są formułowane na podstawie własnej opinii"[58].

Niestety, zaprzęgnięcie algorytmów do rozpoznawania twarzy nie rozwiązuje tego problemu, i jest to jeden z najlepszych powodów, dla których trzeba z wielką ostrożnością podchodzić do stosowania

komputerów w identyfikacji przestępców. Podobieństwo i identyczność nie są i nigdy nie będą tym samym, bez względu na to, jak dobre będziemy mieć algorytmy.

Jest jeszcze powód, żeby uważać, gdzie stawiamy stopy na ścieżce komputeryzacji procesu rozpoznawania twarzy. Tak się bowiem składa, że algorytmy wcale nie radzą sobie z tym aż tak dobrze, jak można by sądzić.

Jedna na milion?

Algorytmy stosują jedną z dwóch podstawowych metod. Pierwsza polega na stworzeniu trójwymiarowego modelu twarzy przez połączenie serii dwuwymiarowych zdjęć albo skanowania twarzy kamerą na podczerwień. Tę metodę stosuje funkcja Face ID w telefonach firmy Apple. Algorytmy z tej grupy zdołały pokonać problemy związane z różnorodnością mimiki i zmianami spowodowanymi starzeniem się dzięki analizie obszarów twarzy kształtowanych przez twardą tkankę i kości, takich jak przebieg krzywizny oczodołu czy grzbietu nosa.

Firma Apple twierdzi, że prawdopodobieństwo odblokowania przez przypadkową osobę telefonu zabezpieczonego funkcją Face ID wynosi jeden do miliona, ale algorytm iPhone'a nie jest bezbłędny – daje się oszukać bliźniakom[59], rodzeństwu[60] i dzieciom właścicieli telefonów. (Wkrótce po inauguracji Face ID w sieci pojawił się filmik, na którym dziesięciolatek odblokowuje wyposażony w algorytm rozpoznawania twarzy telefon swojej matki. Teraz kobieta musi kasować swoje esemesy, jeżeli chce coś ukryć przed synem[61]). Były też doniesienia, że iPhone'a można oszukać za pomocą wydrukowanej na drukarce 3D specjalnej maski z naklejonymi w miejscu oczu dwuwymiarowymi zdjęciami wykonanymi kamerą na światło podczerwone[62]. Wszystko to oznacza, że choć algorytm rozpoznawania twarzy może być już na tyle dobry, żeby skutecznie blokować twój telefon, prawdopodobnie nie jest jeszcze dość wiarygodny, by strzec dostępu do twoich rachunków bankowych.

„Trójwymiarowe" algorytmy nie nadają się również za bardzo do skanowania zdjęć paszportowych ani analizowania zdjęć z kamer

monitoringu. Do tego celu potrzeba już algorytmów innego rodzaju, które trzymają się dwóch wymiarów i wykorzystują statystykę. Nie interesują ich charakterystyczne cechy twarzy, które ludzie mogliby uznać za znaki szczególne; ich zadaniem jest tworzenie statystycznego opisu wzorców plam światła i cienia na zdjęciu – podobnie jak to było w przypadku algorytmów do rozpoznawania psów, o których pisałam w rozdziale *Medycyna*. Całkiem niedawno badacze uzmysłowili sobie, że to wcale nie ludzie muszą decydować o tym, które wzorce sprawdzą się najlepiej (w rozpoznawaniu twarzy) – można zaprogramować algorytm tak, żeby pracując na dostatecznie obszernej bazie danych, metodą prób i błędów sam tworzył najlepsze kombinacje. Najczęściej zaprzęga się do tej pracy sieci neuronowe. I to właśnie tego rodzaju algorytmy zrobiły ostatnio wielki krok naprzód w jakości i skuteczności działania. Ma to jednak swoją cenę: nie zawsze wiemy dokładnie, n a j a k i e j p o d s t a w i e algorytm decyduje o podobieństwie dwóch twarzy.

To zaś oznacza, że te najnowocześniejsze algorytmy można też bardzo łatwo przechytrzyć. Ponieważ wykorzystują one opisy statystyczne wzorców światła i cienia na twarzy, możesz je oszukać, zakładając śmieszne okulary w ramkach zadrukowanych wymyślnymi wzorami. Albo jeszcze lepiej: zaprojektuj specjalny wzór, który będzie „zastępował" czyjąś twarz, a zdołasz wręcz wmówić algorytmowi, że jesteś kimś innym – tak właśnie zrobił gość na poniższym zdjęciu, który dzięki takim okularom „stał się" dla algorytmu aktorką Millą Jovovich[63].

Abstrahując od ataków facetów w śmiesznych okularach, możliwości rozpoznawania twarzy demonstrowane przez algorytmy statystyczne wywołały falę niekłamanego zachwytu mediów. Jednym z takich algorytmów był FaceNet firmy Google. Aby przetestować jego zdolności, podsunięto mu do rozpoznania 5000 zdjęć celebrytów. Ludzie, którym wcześniej dano to samo zadanie, wypadli nadzwyczaj dobrze, osiągając 97,5 procent poprawności (co chyba nie dziwi, przecież twarze celebrytów były im wcześniej dobrze znane)[64]. Ale FaceNet poradził sobie jeszcze lepiej, osiągając fenomenalny wynik 99,6 procent poprawnych rozpoznań.

Wydawać by się mogło, że maszyny opanowały sztukę rozpoznawania w stopniu nieosiągalnym dla człowieka. Wynik robi takie wrażenie, że dla niektórych stanowiłby wystarczającą gwarancję jakości w rozpoznawaniu przestępców. Pojawia się tu jednak jedno małe „ale". Pięć tysięcy twarzy to w rzeczywistości żałośnie mały poligon dla algorytmu. Jeśli komputer ma zostać zaprzęgnięty do walki z przestępczością, będzie musiał umieć rozpoznać jedną twarz spośród nie tysięcy, ale milionów.

Brytyjska policja ma obecnie bazę 19 milionów zdjęć twarzy, na którą składają się wszystkie fotografie wykonane osobom podejrzanym o popełnienie przestępstwa. FBI dysponuje bazą liczącą 411 milionów zdjęć, która obejmuje podobno połowę wszystkich dorosłych Amerykanów[65], a w Chinach, gdzie baza danych, złożona z fotografii z dokumentów tożsamości, daje łatwy dostęp do miliardów twarzy, władze już zainwestowały grube pieniądze w ich rozpoznawanie: na ulicach, stacjach metra i lotniskach zainstalowano kamery, które podobno rejestrują wszystkich, od poszukiwanych przestępców po niefrasobliwych pieszych[66]. (Pojawiły się nawet domysły, że drobne naruszenia porządku, takie jak zaśmiecanie ulic, wpływają na ogólny wynik w punktacji Sesame Credit i powodują wymierzanie kar, o których mówiliśmy w rozdziale *Dane*).

W tym cały problem: ryzyko błędnego rozpoznania rośnie w zastraszającym tempie wraz ze wzrostem ilości zdjęć zgromadzonych

w bazie danych. Im więcej twarzy algorytm ma do porównania, tym większe prawdopodobieństwo, że dwie z nich uzna za podobne. A zatem jeżeli zaczniesz używać algorytmów statystycznych do analizy większych katalogów zdjęć, ich dokładność drastycznie spadnie.

To trochę tak, jakbym bezbłędnie przypisała dziesięć dowodów osobistych do dziesięciu obcych mi osób, po czym – mając dowiedzioną stuprocentową dokładność w rozpoznawaniu twarzy – dostałabym zadanie polegające na chodzeniu po centrum Nowego Jorku i wyłapywaniu w tłumie twarzy poszukiwanych zbrodniarzy. Oczywiście moja dokładność musiałaby spaść znacząco.

Zupełnie tak samo jest z algorytmami. W 2015 roku University of Washington zorganizował konkurs pod hasłem MegaFace, w którym ludzie z całego świata mogli przetestować swoje algorytmy rozpoznawania twarzy na bazie danych liczącej milion zdjęć[67]. To ciągle znacznie mniej niż katalogi niektórych służb, ale to już coś. Niestety, algorytmy nie wypadły za dobrze.

FaceNet firmy Google – który był bliski doskonałości w rozpoznawaniu celebrytów – nagle potrafił poprawnie rozpoznać zaledwie 75 procent[22*] twarzy[68]. Pozostałe algorytmy wypadły naprawdę żałośnie, osiągając zaledwie 10 procent poprawności. Kiedy piszę tę książkę, rekordzistą świata jest chiński algorytm zwany Tencent YouTu Lab, który rozpoznaje twarze z trafnością równą 83,29 procent[69].

Można wyrazić ten wynik inaczej: jeżeli szukasz konkretnego przestępcy w cyfrowej bazie danych liczącej miliony zdjęć, n a j l e p s z y algorytm przeoczy poszukiwaną twarz raz na sześć wyszukiwań.

Powinnam przy tym dodać, że postępy w tej dziedzinie są bardzo szybkie. Wyniki algorytmów są coraz lepsze i nikt nie potrafi powiedzieć na pewno, co się wydarzy w ciągu paru lat czy nawet kilku miesięcy. Ja wiem jednak, że różnice w oświetleniu, pozycji ciała, jakości zdjęcia i ogólnym wyglądzie szalenie utrudniają dokładne i wiarygodne rozpoznanie twarzy przez komputer. Przed nami jeszcze długa droga do tego, żeby uzyskać idealną dokładność na podstawie baz danych liczących 411 milionów zdjęć czy znaleźć sobowtóra – jednego z biliona.

Szukanie równowagi

Po tym kuble zimnej wody mam coś na pocieszenie. Istnieją algorytmy rozpoznawania twarzy, które nadają się do wykorzystania w pewnych określonych sytuacjach. Na przykład w kanadyjskim Ontario osoby uzależnione od hazardu mogą zapisać się dobrowolnie do systemu, który uniemożliwia im wchodzenie do kasyn gier; jeżeli mimo postanowienia ulegną nałogowi, algorytm rozpoznawania twarzy zasygnalizuje personelowi, że trzeba je uprzejmie wyprosić z lokalu[70]. Oczywiście to niesprawiedliwe wobec osób, które na podstawie fałszywego rozpoznania nie będą mogły się rozerwać przy ruletce, ale to chyba cena, jaką warto zapłacić, żeby pomóc człowiekowi wychodzącemu z nałogu oprzeć się pokusie powrotu do hazardu.

Podobnie jest w handlu. Dawniej ściany pomieszczeń dla ochroniarzy były wyklejone zdjęciami sklepowych złodziei, dziś algorytmy rozpoznawania twarzy sprawdzają, czy nie figurujesz w bazie notowanych kleptomanów, gdy tylko wejdziesz do sklepu. Jeżeli zdjęcie twojej twarzy zostanie dopasowane do któregoś z bazy, ochroniarze dostaną powiadomienie na smartfony i pójdą cię szukać między półkami.

Są ważne powody, dla których sklepy chcą korzystać z takich urządzeń. Ocenia się, że w samej Wielkiej Brytanii dochodzi corocznie w sklepach do 3,6 miliona kradzieży, co przynosi właścicielom zatrważające straty – w wysokości 660 milionów funtów[71]. A jeśli weźmie się pod uwagę, że w 2016 roku w amerykańskich sklepach odnotowano 91 przypadków gwałtownych zgonów osób podejrzanych o kradzież[72], bez trudu da się obronić tezę, że stosowanie metod pozwalających na zatrzymanie notowanego złodzieja przed wejściem do sklepu, zanim dojdzie do niebezpiecznej sytuacji, jest w naszym wspólnym interesie.

Zaawansowana technika zapobiegania kradzieżom w sklepach ma jednak swoje minusy: po pierwsze, cierpi na tym nasza prywatność (firma FaceFirst, jeden z największych producentów tego rodzaju programów dla handlu, twierdzi, że nie przechowuje zdjęć stałych klientów, ale sklepy z pewnością korzystają z techniki rozpoznawania twarzy do śledzenia naszych nawyków zakupowych). No i musi paść pytanie o to, kto trafia na taką cyfrową czarną listę. Czy na pewno wszystkie figurujące na niej osoby

znalazły się tam z właściwego powodu? A co z prawnym domniemaniem niewinności? A co z osobami, które znalazły się na liście przypadkowo – w jaki sposób mogą z niej „zniknąć”? Poza tym wciąż istnieje potencjalne ryzyko fałszywego rozpoznania przez algorytm, który nigdy nie będzie stuprocentowo nieomylny.

Musimy ocenić, czy plusy przeważają nad minusami. Odpowiedź na to pytanie nie jest wcale łatwa. Nie zgadzają się w tej kwestii nawet sami właściciele sklepów. Jedni entuzjastycznie montują nowe urządzenia, drudzy znów zaczynają się wycofywać – między innymi Walmart, który zakończył pierwszy test FaceFirsta w marketach sieci, kiedy zwrot inwestycji okazał się mniejszy od zakładanego[73].

Niemniej w przypadku przestępstw ustalenie bilansu szkód i korzyści jest znacznie łatwiejsze – mimo że nie tylko algorytmy stosowane w kryminalistyce bazują na dość kruchej podstawie statystycznej. Nie znamy poziomu błędu także dla badań daktyloskopijnych[74], analizy śladów zębów, analizy śladów krwawych[75] ani balistyki[76]. Jeżeli wierzyć tezom z artykułu opublikowanego w 2009 roku przez Narodową Akademię Nauk Stanów Zjednoczonych, żadna z technik kryminalistycznych poza badaniami genetycznymi nie pozwala na „powiązanie materiału dowodowego z konkretną osobą albo źródłem”[77]. Mimo to nikt nie zaprzeczy, że wszystkie stanowią niezwykle pomocne narzędzia w pracy policyjnej – pod warunkiem że nie będziemy bezgranicznie wierzyć we wnioski, jakie pozwalają wyciągać. No, ale dokładność nawet najnowocześniejszych algorytmów rozpoznawania twarzy pozostawia jeszcze bardzo wiele do życzenia. Pojawiły się wręcz głosy, że jeżeli istnieje choćby najmniejsze ryzyko powtórzenia się historii Steve’a Talleya, to nie powinniśmy korzystać z niedoskonałej techniki, aby przypadkiem nie pozbawić kogoś niesłusznie wolności. Kłopot jednak w tym, że takie historie nie wyczerpują tematu. Sprawa wygląda bowiem tak, że choć używanie algorytmów rozpoznających twarze do łapania przestępców ma ogromne wady, oferuje nam jednocześnie niebagatelne korzyści.

Ryzykowny barter

W maju 2015 roku ulicami Manhattanu biegał mężczyzna i atakował przypadkowych przechodniów czarnym młotkiem ciesielskim. Najpierw podbiegł do grupki osób przy Empire State Building i uderzył dwudziestolatka w tył głowy. Po sześciu godzinach pojawił się na Union Square i tym samym młotkiem uderzył w skroń kobietę siedzącą spokojnie na ławce w parku. Zaledwie kilka minut później znów zaatakował, tym razem obrawszy sobie za cel 33-letnią kobietę, która szła ulicą na wysokości parku[78]. Na podstawie zdjęć miejsc napaści pobranych z kamer monitoringu algorytm rozpoznawania twarzy zidentyfikował napastnika – był to David Baril, który kilka miesięcy przed atakami zamieścił na Instagramie zdjęcie młotka ociekającego krwią[79]. Przyznał się do postawionych mu zarzutów i został skazany na karę 22 lat pozbawienia wolności.

Dzięki przełomowi zapoczątkowanemu przez algorytmy rozpoznawania twarzy policja wznawia dawno zarzucone sprawy. W 2014 roku algorytm doprowadził do postawienia przed sądem Amerykanina, który przez piętnaście lat ukrywał się za granicą pod fałszywym nazwiskiem. Neil Stammer, oskarżony między innymi o molestowanie i porwanie nieletnich, uciekł po zwolnieniu z aresztu za poręczeniem majątkowym; został ponownie aresztowany, kiedy algorytm porównał jego zdjęcie z listu gończego FBI z fotografiami z bazy danych paszportowych i dopasował je do zdjęcia mieszkańca Nepalu, który nosił paszport wydany na inne nazwisko[80].

Jak wiele mogą znaczyć algorytmy rozpoznawania twarzy, wiemy od lata 2017 roku, kiedy osiem osób zginęło w ataku terrorystycznym na London Bridge. Youssef Zaghba był jednym z trzech mężczyzn, którzy wjechali vanem w pieszych, a potem zaczęli atakować ludzi nożami na sąsiednim Borough Market. Zaghba znajdował się na liście osób podejrzanych o terroryzm we Włoszech i przed wjazdem do Wielkiej Brytanii mógłby zostać automatycznie zidentyfikowany przez algorytm rozpoznawania twarzy.

Czy rzeczywiście musielibyśmy wymienić własną prywatność na ochronę, a sprawiedliwość na bezpieczeństwo? Ilu Steve'ów Talleyów

jesteśmy gotowi poświęcić w zamian za szybkie ujęcie takich ludzi jak David Baril i Youssef Zaghba?

Przyjrzyjmy się statystykom nowojorskiej policji. W 2015 roku skutecznie rozpoznano 1700 podejrzanych, z których 900 aresztowano; błędnych rozpoznań odnotowano pięć[81]. I choć za każdym z tych pięciu fałszywych rozpoznań stoi przykra historia, pozostaje pytanie: czy to akceptowalny margines błędu? Czy to cena, jaką jesteśmy gotowi zapłacić za ograniczenie przestępczości?

Okazuje się, że algorytmy pozbawione wad, takie jak algorytm profilowania geograficznego Kima Rossma, o którym mówiliśmy na początku tego rozdziału, są raczej wyjątkiem niż regułą. W każdej dziedzinie zwalczania przestępczości znajdziemy algorytmy, które z jednej strony obiecują nam bardzo wiele, a z drugiej – potrafią nam spędzać sen z powiek. PredPol, HunchLab, Strategic Subject List i algorytmy rozpoznawania twarzy – wszystkie mają być panaceum na nasze zmartwienia, a każdy z osobna przysparza nam po drodze nowych kłopotów.

W żadnej dziedzinie wykorzystanie algorytmów nie wymaga tak oczywistej i naglącej regulacji jak właśnie w dziedzinie zapobiegania przestępczości, gdzie już samo istnienie systemów opartych na algorytmach rodzi poważne pytania, na które nie mamy łatwych odpowiedzi. Będziemy musieli się zmierzyć z tymi dylematami. Czy powinniśmy domagać się stosowania tylko tych algorytmów, które potrafimy zrozumieć i rozebrać na czynniki pierwsze – ze świadomością, że odebranie ich właścicielom może oznaczać zmniejszenie ich skuteczności (a więc także zwiększenie się liczby przestępstw)? Czy zdyskwalifikujemy systemy matematyczne z ich nieuchronnymi uprzedzeniami i doświadczalnie potwierdzoną omylnością – wiedząc, że będziemy stawiać im tym samym wyższe wymagania niż systemom tradycyjnym, z którymi zostaniemy? I jeszcze: kiedy uprzedzenia przestają być akceptowalne? W którym momencie ważniejsze od ofiar algorytmów stają się ofiary przestępstw, którym możemy zapobiegać?

W pewnym sensie wszystkie te odpowiedzi zależą od tego, co jako społeczeństwo uznamy za sukces. Jakie są nasze priorytety? Czy chodzi o to, żeby maksymalnie ograniczyć przestępczość? Czy raczej postawić na

pierwszym miejscu nietykalność niewinnych? Ile jednego bylibyśmy gotowi poświęcić dla uzyskania drugiego?

Gary Marx, profesor socjologii w amerykańskim MIT, doskonale ujął ten dylemat w wywiadzie dla „Guardiana": „Związek Sowiecki praktycznie nie znał zjawiska drobnej przestępczości ulicznej, kiedy znajdował się w swojej najgorszej fazie totalitaryzmu. Tylko że, mój Boże – za jaką cenę?"[82].

Niewykluczone, że ostatecznie dojdziemy do wniosku, iż algorytmom należy wyznaczyć pewne granice. Że pewnych rzeczy nie należy analizować ani obliczać. I niewykluczone, że ta sentymentalna postawa obejmie także inne dziedziny życia poza światem zbrodni. I chyba nie stanie się tak z powodu braku dobrych chęci ze strony algorytmów, ale dlatego że być może jednak są pewne rzeczy, których beznamiętna maszyna nie jest w stanie zrozumieć.

Sztuka

Czwartego lutego 2018 roku Justin był w refleksyjnym nastroju. Siedział na kanapie w salonie swojego domu w Memphis w stanie Tennessee i oglądał relację z meczu Super Bowl, podjadając M&M's-y. Parę dni wcześniej obchodził swoje 37. urodziny i zgodnie z coroczną tradycją dumał teraz nad swoim życiem.

Wiedział, że tak naprawdę powinien się cieszyć. Żył sobie jak pączek w maśle. Regularna praca w biurze od dziewiątej do piątej, dach nad głową i kochająca rodzina. A przecież zawsze chciał czegoś więcej. Jako nastolatek wierzył, że są mu pisane pieniądze i sława.

Jak więc to się stało, że skończył tak… normalnie? „Wszystko przez ten boys band", pomyślał. Ten, do którego dołączył, jak miał 14 lat. „Gdyby się nam udało, moje życie wyglądałoby teraz zupełnie inaczej". Ale nie wiedzieć czemu zespół okazał się niewypałem. Sukces jakoś nie chciał zapukać do drzwi biednego Justina Timberlake'a.

Przygnębiony Justin otworzył kolejne piwo i zaczął snuć marzenia o sławie, która nigdy nie nadeszła.

Skończyły się reklamy, rozpoczął się koncert w przerwie meczu. A w świecie równoległym – identycznym z tym poza jednym drobnym szczegółem – na scenę wszedł właśnie inny 37-letni Justin Timberlake z Memphis w stanie Tennessee.

Światy równoległe

Dlaczego prawdziwy Justin Timberlake odniósł tak wielki sukces? I dlaczego innemu Justinowi Timberlake'owi już się nie udało? Niektórzy (w tym ja w wieku 14 lat)[23*] upieraliby się, że sukces gwiazdy muzyki pop Justina Timberlake'a jest zasłużony: kiedy ktoś ma wrodzony talent, świetną prezencję, umiejętności taneczne i pisze utwory o przyzwoitej wartości artystycznej, sława jest nieunikniona. Znaleźliby się jednak i tacy,

którzy nie chcieliby się z tym zgodzić. Mogliby twierdzić, że Justin Timberlake nie jest wcale wyjątkowy, podobnie zresztą jak żadna z wielbionych przez rzesze fanów gwiazd muzyki pop. Odkryć talent wokalny i taneczny jest bardzo łatwo – gwiazdy po prostu miały więcej szczęścia od innych.

Oczywiście nie sposób stwierdzić tego na pewno. Trzeba by skonstruować mnóstwo identycznych światów równoległych, do każdego wprowadzić takiego Timberlake'a, po czym obserwować rozwój każdego awatara, żeby zobaczyć, czy wszędzie odniesie podobny sukces, lecz niestety, stworzenie takiego kompleksu sztucznych światów przekracza nasze możliwości. Jeżeli jednak obniżymy sobie trochę poprzeczkę i weźmiemy pod lupę mniej znanych artystów, jesteśmy w stanie zbadać względną rolę szczęścia i talentu w notowaniach utworów muzycznych.

Właśnie ta idea stała się podstawą słynnego eksperymentu z 2006 roku. Jego twórcy – Matthew Salganik, Peter Dodds i Duncan Watts – zbudowali kilka cyfrowych światów[1]. Najpierw zaprojektowali własny internetowy odtwarzacz muzyki, taką prymitywną wersję Spotify, i przypisali jego użytkowników do ośmiu równoległych portali muzycznych, na których można było znaleźć ten sam zestaw 48 utworów nieznanych wykonawców.

W ramach projektu, który nazwano MusicLab[2], do odtwarzacza zalogowało się w sumie 14 341 fanów muzyki; każdy miał dostęp do wszystkich utworów, mógł je oceniać i ściągać ulubione kawałki na swoje konto.

Podobnie jak w prawdziwym Spotify, użytkownicy mieli ciągły podgląd tego, czego słuchają inni w ich „świecie". Przy nazwisku artysty lub nazwie zespołu i tytule piosenki wyświetlała się aktualna lista ściągnięć utworu – oczywiście także w ramach „ich świata". Początkowo wszystkie liczniki były wyzerowane, ale z czasem liczby się zmieniały i stopniowo zaczęli się uwidaczniać liderzy ośmiu równoległych list przebojów.

W celu zmierzenia naturalnej, „prawdziwej" popularności utworów badacze skonstruowali również świat kontrolny, gdzie popularność piosenek nie mogła wpływać na wybór jednej z nich przez użytkowników. W tym świecie były one wyświetlane na stronie w przypadkowej

kolejności – w formie diagramu albo listy – a statystyki ściągnięć zostały ukryte.

Wyniki były intrygujące. We wszystkich światach zgodnie uznano niektóre utwory za oczywisty chłam. Inne znów stały się wyraźnymi faworytami – zyskały dużą popularność w każdym świecie, także w tym kontrolnym, gdzie nie było widać liczby ściągnięć. Jednak w przedziale między pewnymi hitami i totalnymi kitami artyści mogli się przekonać na własnej skórze, że fortuna kołem się toczy.

Weźmy dla przykładu 52Metro, punkową kapelę z Milwaukee, której piosenka *Lockdown* cieszyła się wielką popularnością w jednym świecie – wspięła się tam na sam szczyt listy przebojów – a w innym znów została kompletnie zignorowana, zajmując zaledwie 40. miejsce na liście 48 utworów. To była dokładnie ta sama piosenka w dokładnie takim samym otoczeniu, ale akurat w tym świecie zespół 52Metro po prostu się nie przyjął[3]. Sukces bywa jednak wypadkową szczęścia.

Mimo że nie istniała pewna recepta na drogę na szczyt listy przebojów, badacze stwierdzili, że prawdopodobieństwo ściągnięcia utworu znacznie wzrastało, kiedy użytkownik wiedział, że piosenka jest lubiana przez innych. Jeżeli całkiem przeciętna kompozycja znalazła się przypadkiem na pierwszym miejscu listy we wczesnej fazie eksperymentu, mogła się stać zdumiewająco popularna; poprzednie ściągnięcia napędzały kolejne. Domniemana popularność zamieniała się w rzeczywistą, a zatem ostateczny sukces był po prostu funkcją przypadku spotęgowaną przez czas.

Jest wytłumaczenie, dlaczego tak się działo. Zjawisko to psychologowie znają pod pojęciem s p o ł e c z n e g o d o w o d u s ł u s z n o ś c i. Kiedy nie mamy dość informacji, żeby samodzielnie podjąć decyzję, nawykowo naśladujemy zachowanie otoczenia. To dlatego teatry umieszczają niekiedy na widowni klakierów, żeby klaskali czy wiwatowali we właściwym momencie. Kiedy słyszymy, że ludzie wokół nas zaczynają bić brawo, jesteśmy bardziej skłonni zareagować tak samo. Z muzyką jest podobnie – może niekoniecznie najbardziej lubimy to, czego słuchają inni, niemniej postawienie na popularność zabezpiecza nas przed rozczarowaniem. „Ludzie mają za dużo opcji – powiedział Salganik w wywiadzie dla «LiveScience». – Nie możesz odsłuchać wszystkich

utworów, więc naturalną koleją rzeczy idziesz na skróty i wybierasz to, czego słuchają inni”[4].

Traktujemy popularność jako wyznacznik jakości we wszystkich dziedzinach rozrywki. W badaniu z 2007 roku naukowcy przeanalizowali wpływ listy bestsellerów gazety „New York Times” na odbiór książki przez opinię publiczną. Autor tego studium, Alan Sorensen, zastosował specyficzne kryteria, jakimi kierowali się twórcy listy „NYT”, i wziął pod lupę sukces rynkowy książek, które na podstawie wyników sprzedaży powinny się znaleźć na liście, lecz z powodu opóźnień premiery czy niezamierzonych przeoczeń na nią nie trafiły, a następnie porównał go z sukcesem książek z listy. Różnice były dramatyczne: już samo znalezienie się na liście „New York Timesa” zwiększało sprzedaż średnio o 13–14 procent, a w przypadku debiutów wzrost wynosił aż 57 procent.

Im więcej źródeł używamy do określenia popularności – takich jak listy bestsellerów, rankingi Amazona, notowania na Rotten Tomatoes, listy Spotify – tym baczniejszą zwracamy uwagę na społeczny dowód słuszności. Ten wpływ potęguje się jeszcze, kiedy zostajemy zarzuceni milionami opcji, bo o naszą uwagę walczą też reklamy, recenzje i media.

Wszystko to oznacza, że czasem na sam szczyt może trafiać muzyka bardzo kiepska. Podobno w latach dziewięćdziesiątych ubiegłego wieku dwaj brytyjscy producenci muzyczni założyli się o to, który zdoła umieścić na listach przebojów najgorszą piosenkę na świecie. To właśnie wskutek tego zakładu miał się na nich pojawić zespół dziewczęcy Vanilla, którego debiut, zatytułowany *No Way No Way*, był przeróbką słynnego kawałka *Mahna Mahna* z *Muppet Show*. Klip pokazywał występ dziewcząt, które tylko osoba obdarzona naprawdę bujną wyobraźnią mogłaby określić mianem piosenkarek, a zaprezentowana w nim grafika wyglądała jak zrobiona pod windowsowym Paintem; z czystym sumieniem można by to zgłosić do konkursu na najgorszy materiał promocyjny, jaki kiedykolwiek pokazano publicznie[5]. Vanilla miała jednak potężnych sprzymierzeńców – dzięki artykułom w kilku czasopismach i występowi w programie muzycznym *Top of the Pops* na antenie BBC piosenka zdołała wspiąć się na 14. miejsce na liście przebojów[24*].

Trzeba przyznać, że sukces zespołu był chwilowy. Kiedy pojawił się drugi singiel, Vanilla zaczęła tracić popularność. Trzeciego nie wydano nigdy. To może oznaczać, że nie wszystko jednak zależy od społecznego dowodu słuszności – co zresztą pokazał kolejny eksperyment twórców MusicLabu.

Jego scenariusz był taki sam jak w pierwszym eksperymencie, tym razem jednak – aby sprawdzić, w jakim stopniu domniemana popularność może być samospełniającą się przepowiednią – badacze wykonali małą woltę. Po ustabilizowaniu się list przebojów we wszystkich światach eksperyment został wstrzymany, a listy odwrócone do góry nogami. Teraz nowi użytkownicy ich odtwarzacza muzyki widzieli najbardziej popularnych wykonawców na samym dole list, a niewypały zajmujące dotychczas doły tabeli przemieniły się nagle w gwiazdy światowej sławy.

Niemal od razu ogólna liczba ściągnięć znacząco spadła. Kiedy na samym szczycie znalazły się utwory nieciekawe, użytkownicy stracili zainteresowanie resztą prezentowanej na stronie muzyki. Największe spadki liczby ściągnięć dotknęły dotychczasowe niewypały, znajdujące się teraz na pierwszych miejscach list. Dobre piosenki okupujące doły tabel radziły sobie gorzej niż wtedy, kiedy zajmowały pierwsze miejsca na listach, ale ciągle lepiej od tych, które poprzednio uplasowały się na miejscach ostatnich. Gdyby badacze odpowiednio przedłużyli tę fazę eksperymentu, najlepsze utwory odzyskałyby dawną popularność. Wniosek: rynek nie jest niezmienny w czasie. Oba czynniki, szczęście i jakość, mają swój udział w sukcesie danego utworu[6].

Tymczasem w naszej rzeczywistości – gdzie mamy tylko jeden świat danych do obróbki – istnieje prosta interpretacja spostrzeżeń płynących z eksperymentu MusicLab. Jakość ma znaczenie i nie jest tożsama z popularnością; najlepsze utwory odzyskały swoje dawne wysokie notowania, a zatem pewne piosenki są po prostu „z natury" lepsze od innych. Fantastyczny utwór napisany przez fantastycznego wykonawcę powinien być (przynajmniej teoretycznie) skazany na sukces. Nie musi to jednak wcale oznaczać, że będzie także odwrotnie: z tego, że utwór jest popularny, nie wynika automatycznie, że jest wysokiej jakości.

Zupełnie osobną sprawą jest to, jak tę jakość zdefiniować – i do tego przejdziemy za chwilę. Nie dla wszystkich jednak jakość jest najważniejsza. Każda wytwórnia fonograficzna, każdy producent filmowy

i każde wydawnictwo chciałyby znać odpowiedź na pytanie za milion dolarów: czy można z góry rozpoznać materiał na pewny sukces? Czy istnieje algorytm zdolny wytypować przebój?

Polowanie na hity

Inwestowanie w produkcję filmową to ryzykowna gra. Niewiele filmów przynosi jakiś zysk, większość co najwyżej zapewnia zwrot poniesionych kosztów; klapa to w branży filmowej szara codzienność[7]. Stawki są tu wysokie: kiedy koszty filmu idą w dziesiątki, czy nawet setki milionów dolarów, skutki błędnej oceny popytu na produkt mogą kosztować dosłownie fortunę.

Doświadczyła tego na własnej skórze wytwórnia Disney z nakręconym w 2012 roku filmem *John Carter*. Wpakowano w jego produkcję 350 milionów dolarów, z niezłomnym przekonaniem, że zajmie poczesne miejsce obok najlepszych Disneyowskich franczyz, *Toy Story* i *Gdzie jest Nemo?* Nie widzieliście tego filmu? Ja też nie. *John Carter* nie zdołał zawojować publiczności i skończył ze stratą 200 milionów dolarów, co doprowadziło do dymisji szefa Walt Disney Studios[8].

Bonzowie Hollywoodu zawsze brali pod uwagę to, że sukcesu komercyjnego filmu nie da się przewidzieć. To dziedzina, w której po prostu trzeba mieć nosa. Jack Valenti, prezes zarządu i CEO stowarzyszenia Motion Picture Association of America, tak to ujął w 1978 roku: „Nikt ci nie powie, jak film poradzi sobie na rynku. To się wie dopiero wtedy, kiedy w ciemnej sali kinowej rusza czołówka i między ekranem a widownią zaczynają przeskakiwać iskry”[9]. Zwięźlej wyraził to pięć lat później William Goldman, autor scenariusza filmów takich jak *Narzeczona księcia* oraz *Butch Cassidy i Sundance Kid*: „Nikt nic nie wie”[10].

Ale w tej książce już nieraz się przekonaliśmy, że nowoczesne algorytmy potrafią dziś rutynowo przewidywać rzeczy jeszcze wczoraj pozornie nieprzewidywalne. Dlaczego przemysł rozrywkowy miałby się różnić od innych dziedzin? Sukces filmu da się zmierzyć – zarówno wysokością przychodów, jak i oceną krytyków. Można zmierzyć wiele elementów struktury i cech filmu: obsadę, gatunek, budżet, czas projekcji, wątki fabuły

i tak dalej. Dlaczego by więc nie spróbować zastosować algorytmu do poszukiwania kur znoszących złote jajka: filmów, które po prostu są skazane na sukces kasowy?

Takie ambitne założenia mieli autorzy kilku współczesnych badań naukowych: postanowili skorzystać z niezwykle bogatych źródeł danych, gromadzonych i udostępnianych przez takie witryny jak Internet Movie Database (IMDb) czy Rotten Tomatoes. Jak się oczywiście okazało, dane te skrywają wiele intrygujących informacji.

Weźmy badanie przeprowadzone w 2013 roku przez Sameeta Sreenivasana[11], który zauważył, że dzięki użytkownikom bazy IMDb, opatrującym fabuły filmów frazami kluczowymi, powstał niewiarygodnie szczegółowy zasób deskryptorów, na którym można zademonstrować, jak ewoluowały w czasie nasze gusta filmowe. Kiedy Sreenivasan zaczynał swoje badania, katalog IMDb liczył ponad dwa miliony filmów nakręconych w ciągu ponad stu lat – każdy był oznaczony wieloma różnymi frazami. Niektóre z nich były bardzo ogólne, jak „przestępczość zorganizowana" albo „relacja ojciec–syn", inne nawiązywały do lokalizacji akcji, na przykład „manhattan-new-york-city", albo do konkretnych elementów fabuły, jak „grożenie bronią" czy „związanie na krześle".

Już same te frazy kluczowe pokazują, że nasze zainteresowanie określonymi elementami fabuły podlega sezonowym wahaniom; wystarczy wspomnieć filmy z czasów drugiej wojny światowej, które często poruszają temat aborcji. Sreenivasan zauważył, że zbliżone tematycznie produkcje pojawiają się w krótkich odstępach czasu, po czym następuje przerwa w eksploatacji wątku. Analiza wszystkich fraz kluczowych pozwoliła wyrazić liczbą nowatorstwo danego filmu w okresie jego pojawienia się na ekranach i porównać tę liczbę z jego sukcesem kasowym.

Jeżeli określony element albo kluczowa cecha fabuły – powiedzmy: kobieca nagość czy zorganizowana przestępczość – były znane z wcześniejszych produkcji, dawały danemu filmowi niską notę za nowatorstwo. Ale każda oryginalna, niespotykana dotąd cecha fabuły – na przykład wprowadzenie sztuk walki do filmów akcji w latach siedemdziesiątych – zapewniała filmowi wysoki wynik.

Okazuje się, że mamy z tym nowatorstwem dość złożone relacje. Na ogół jest tak, że im wyższy wynik w tej kategorii miał dany film, tym lepszą miał

oglądalność. Ale tylko do pewnego progu, bo dalej zieje przepaść – wpływy za filmy, które przekroczyły granicę 0,8 punktu za nowatorstwo, spadały dramatycznie. Studium Sreenivasana potwierdziło przypuszczenia socjologów: banał nas mierzi, ale nie cierpimy także radykalnej inności. Najbardziej kasowe firmy rozsiadły się na wąskiej grzędzie między „nowym" a „nie za bardzo nowym".

Wynik za nowatorstwo mógłby stanowić użyteczne narzędzie, pozwalające studiom filmowym unikać topienia pieniędzy w totalnych klapach, nie na wiele się jednak zda, jeżeli chciałbyś poznać przyszłe losy konkretnego filmu; w takim przypadku powinno się skorzystać raczej z prac europejskiego zespołu badawczego. Odkrył on mianowicie związek między liczbą edycji strony filmu w Wikipedii w miesiącu poprzedzającym premierę kinową a późniejszą jego oglądalnością[12]. Edytorami były często osoby niezwiązane ze studiem filmowym – typowi fani kina, którzy dokładali od siebie kolejne informacje o filmie. Więcej edycji przekłada się na większy szum wokół zbliżającej się premiery, a to z kolei na dłuższe kolejki do kasy.

Ten model ma przyzwoitą moc prognostyczną: z 312 filmów uwzględnionych w studium badacze trafnie, z co najmniej 70-procentową dokładnością, przewidzieli dochód z 70 z nich. Im lepiej radził sobie film i im więcej było edycji na jego stronie w Wikipedii, tym więcej danych mieli do dyspozycji badacze i tym dokładniejsze mogły być ich prognozy. Wpływy ze sprzedaży biletów sześciu największych hitów kasowych zostały przewidziane z dokładnością do 99 procent.

Wyniki tych badań są interesujące, ale model, który sprawdza się dopiero na miesiąc przed premierą filmu, nie za bardzo się przyda inwestorom. Może więc trzeba chwycić byka za rogi i po prostu wziąć pod uwagę wszystkie czynniki znane przed premierą – czyli gatunek, popularność pierwszoplanowych aktorów z obsady, kategorię wiekową filmu (PG, 12 i tak dalej) – po czym wrzucić je do algorytmu uczenia maszynowego i kazać mu przewidzieć, czy film okaże się hitem.

W słynnym badaniu z 2005 roku tak właśnie zrobiono: kazano sieci neuronowej przewidzieć popularność filmów na długo przed ich premierą kinową[13]. Aby maksymalnie uprościć sprawę, badacze zrezygnowali z dokładnego przewidywania wpływów kasowych i zdecydowali się

przypisać filmy do jednej z dziewięciu kategorii – od kompletnej klapy do światowego hitu. Niestety, mimo tych uproszczeń rezultaty pozostawiały bardzo wiele do życzenia. Użyta w badaniu sieć neuronowa pobiła co prawda wcześniej stosowane techniki statystyczne, ale zdołała trafnie określić popularność filmu średnio w 36,9 procent przypadków. Wynik był tylko trochę lepszy w najwyższej kategorii – zaliczono do niej filmy przynoszące ponad 200 milionów dochodu – gdzie odsetek poprawnie rozpoznanych hitów wyniósł 47,3 procent. Słówko ostrzeżenia dla inwestorów: mniej więcej 10 procent filmów, które algorytm uznał za pewne hity, zarobiło niecałe 20 milionów, czyli według miar hollywoodzkich tyle co nic.

W kolejnych badaniach autorzy starali się poprawić skuteczność prognozowania, nie doszło jednak na razie do znaczącego przełomu. Wniosek jest jeden: dopóki nie pojawią się informacje o pierwszych reakcjach publiczności, dopóty popularności filmu w zasadzie nie da się przewidzieć. Jeśli musisz z góry wytypować hit z listy produkowanych filmów, powinnaś przyznać rację Williamowi Goldmanowi: nikt nic nie wie.

Kwantyfikacja jakości

Tak więc przewidywanie popularności to bardzo trudna sprawa. Nie istnieje prosta metoda pozwalająca oddzielić to, co wszyscy lubimy, od tego, dlaczego to lubimy. A jest to jednak problem dla algorytmów działających w świecie kreacji. Bo jeżeli nie możemy użyć popularności do określenia, co jest „dobre", to w jaki sposób zmierzymy jakość?

To ważne: jeżeli chcemy, żeby algorytmy miały jakąkolwiek autonomię w dziedzinie sztuki – żeby samodzielnie tworzyć nowe dzieła albo sensownie analizować sztukę tworzoną przez człowieka – będziemy musieli wypracować jakąś miarę jakości. Musi istnieć obiektywna metoda nakierowania algorytmu na właściwy tor, jakaś empiryczna podstawa, na której maszyna będzie mogła zawsze się oprzeć – artystyczny odpowiednik stwierdzeń: „to skupisko komórek jest rakotwórcze" albo „oskarżony później popełnił przestępstwo". Bez tego jakikolwiek postęp nie będzie możliwy. Nie możemy zaprojektować algorytmu tak, żeby komponował

albo wyszukiwał „dobre" piosenki, jeśli nie potrafimy zdefiniować, co to „dobre" oznacza.

Niestety, pracując nad obiektywną miarą jakości, musimy się zmierzyć z bardzo kontrowersyjnym problemem filozoficznym, na który zwracał już uwagę Platon i który od ponad dwóch tysięcy lat nie przestaje być przedmiotem zaciętych sporów. Jak oceniać wartość artystyczną sztuki?

Niektórzy filozofowie, tacy jak Gottfried Leibniz, twierdzili, że skoro pewne wytwory sztuki wszyscy uznajemy za piękne, na przykład *Dawida* Michała Anioła czy *Lacrimosę* Mozarta, to musi istnieć definiowalna i mierzalna esencja piękna, która pozwala obiektywnie stwierdzić, że dane dzieło sztuki jest lepsze od innego.

Rzadko się jednak tak zdarza, żeby wszyscy zgadzali się ze sobą. Inni filozofowie, jak na przykład David Hume, dowodzili, że piękno kryje się w oku patrzącego. Przykładem może być słynne dzieło Andy'ego Warhola, które dla jednych odbiorców jest źródłem głębokich przeżyć estetycznych, dla drugich znów pod względem artystycznym niczym się nie różni od zwykłej puszki zupy.

Jeszcze inni – należał do nich Immanuel Kant – byli przekonani, że prawda leży gdzieś pośrodku. Nasze sądy o pięknie nie są do końca subiektywne, lecz nie mogą być też i całkiem obiektywne, gdyż są zarazem zmysłowe, emocjonalne i umysłowe, a do tego jeszcze mogą się zmieniać w czasie w zależności od stanu umysłu obserwatora.

Mamy trochę dowodów na poparcie tej tezy. Fani Banksy'ego mogą pamiętać, że w 2013 roku ustawił kramik w nowojorskim Central Parku i anonimowo sprzedawał swoje oryginalne dzieła – namalowane sprayem czarno-białe obrazy na tkaninach – po 60 dolarów od sztuki. Kramik stał w rzędzie podobnych stoisk z typowymi gadżetami dla turystów, tak więc cena musiała się wydać wygórowana potencjalnym nabywcom. Minęło wiele godzin, zanim ktoś wreszcie zdecydował się coś od niego kupić. W sumie Banksy zarobił tego dnia 420 dolarów[14]. Ale rok później, w londyńskim domu aukcyjnym, inny kupiec uznał, iż wartość artystyczna tych samych tkanin jest tak wysoka, że za tylko jedną z nich był gotów zapłacić 68 tysięcy funtów (czyli wówczas około 115 tysięcy dolarów)[15].

Oczywiście nie każdy lubi Banksy'ego. Charlie Brooker – twórca *Czarnego lustra* – nazwał go raz „palantem, którego prace wydają się porażająco przenikliwe tylko idiotom"[16]. Można by uznać, że to dowód, iż prace Banksy'ego są pozbawione autentycznej wartości – astronomiczne sumy, jakie osiągają jego dzieła, to wyłącznie wynik szumu medialnego (i społecznego dowodu słuszności). Okazuje się jednak, że kapryśność naszych gustów estetycznych zaobserwowano także w odniesieniu do form artystycznych o bezsprzecznie wysokiej jakości.

Mój ulubiony przykład pochodzi z eksperymentu przeprowadzonego w 2007 roku przez „Washington Post"[17]. Redakcja gazety uprosiła światowej sławy skrzypka Joshuę Bella, żeby do swojego napiętego harmonogramu wyprzedanych koncertów filharmonicznych dodał mały koncert kameralny. Uzbrojony w stradivariusa o wartości 3,5 miliona dolarów muzyk stanął przy ruchomych schodach na jednej ze stacji waszyngtońskiego metra podczas porannego szczytu, położył na ziemi kapelusz i zaczął grać. Grał przez 43 minuty. „Jeden z najwybitniejszych na świecie wykonawców muzyki klasycznej, na jednym z najdroższych instrumentów, jakie kiedykolwiek wyprodukowano, grał najpiękniejszą muzykę, jaką kiedykolwiek skomponowano", napisał „Washington Post". Wynik? Siedem osób przystanęło na chwilę, żeby go posłuchać. Ponad tysiąc innych minęło muzyka bez zatrzymywania się. Po skończonym koncercie Bell miał w kapeluszu marne 32 dolary i 17 centów.

Zmienia się także to, co uznajemy za „dobre". Apetyt na określone gatunki muzyki klasycznej jest zdumiewająco odporny na upływ czasu, tego samego nie można już jednak powiedzieć o innych formach sztuki. Armand Leroi, profesor biologii ewolucyjnej na Imperial College London, przebadał rozwój muzyki pop i w swojej analizie przedstawił wyraźne dowody na to, że nasze gusta się zmieniają. „Populacja ma wbudowany określony próg znudzenia. Narasta napięcie i w końcu ludzie odczuwają potrzebę czegoś nowego"[18].

Jako przykład można podać automaty perkusyjne i syntezatory, które rozpowszechniły się w muzyce pop pod koniec lat osiemdziesiątych ubiegłego wieku – moda na nie była tak wielka, że listy przebojów drastycznie się zhomogenizowały. „Wszystko nagle brzmi jak wczesna Madonna albo jeden z przebojów Duran Duran – wyjaśnia Leroi. –

Co może cię skłonić do wyciągnięcia wniosku, że to już szczyt muzyki pop. Że właśnie o to chodzi, iż znaleźliśmy jego najwyższą formę. Ale tak oczywiście nie jest. Chwilę później pojawia się hip-hop i znowu mamy eksplozję różnorodności w muzyce na listach przebojów". Kiedy spytałam, czy hip-hop miał w sobie coś wyjątkowego, że zaszła tak gwałtowna zmiana, Leroi odpowiedział: „Nie sądzę. To mogło być cokolwiek, ale tak się złożyło, że był to akurat hip-hop. Na który amerykański odbiorca zareagował stwierdzeniem: Wow, to jest coś nowego, chcemy tego więcej".

W tym właśnie sęk. Nawet jeżeli istnieją obiektywne kryteria pozwalające uznać jedno dzieło sztuki za lepsze od innych, to dopóki w naszym estetycznym odbiorze sztuki rolę odgrywa kontekst, dopóty nie da się stworzyć czytelnej miary jakości estetycznej, która będzie działać zawsze i wszędzie. Bez względu na to, jakich użyjesz technik statystycznych, jak inteligentnej sztucznej inteligencji i jakich algorytmów uczenia maszynowego, próby wykorzystania liczb do pomiaru stopnia artystycznej doskonałości przypominają łapanie dymu.

No, ale algorytm m u s i mieć jakiś punkt zaczepienia. Kiedy wykreślisz z równania popularność i niewymierną jakość, pozostaje ci tylko jedna rzecz, którą da się skwantyfikować: podobieństwo do tego, co już zostało stworzone.

Dzięki mierze podobieństwa możemy zdziałać bardzo wiele, a zdaje się, że dla algorytmów rekomendacji – używanych na przykład przez Netflix czy Spotify – jest to w ogóle metoda idealna. Obie platformy pomagają swoim użytkownikom odkrywać nowe filmy i utwory muzyczne, a w ramach subskrypcji – trafnie odczytywać ich upodobania. Nie mogą bazować na tym, co jest akurat popularne, bo użytkownicy byliby bombardowani samymi Justinami Bieberami i filmami kinowymi ze świnką Peppą. Nie mogą też opierać się na opiniach orędowników kultury wysokiej, takich jak krytycy filmowi, bo profile użytkowników byłyby zasypywane nudami z klubów dyskusyjnych, gdy oni po długim dniu pracy chcą po prostu rozłożyć się na kanapie i dać się wciągnąć w beznadziejny horror albo pogapić się dwie godziny na Ryana Goslinga.

Podobieństwo pozwala algorytmowi skoncentrować się wyłącznie na preferencjach użytkowników. Czego lubią słuchać, co oglądać, do czego wracać od czasu do czasu? Z tymi danymi wystarczy zajrzeć do IMDb

i Wikipedii, na blogi muzyczne albo do artykułów w czasopismach i wygenerować ciąg odpowiednich fraz kluczowych dla każdej piosenki, każdego artysty czy filmu. Po przejrzeniu całego katalogu wystarczy już tylko odszukać i zaproponować inne utwory i filmy opatrzone podobnymi wyrażeniami kluczowymi. Dodatkowo można zajrzeć na profile innych użytkowników, którzy lubią podobne filmy i utwory, żeby sprawdzić, co jeszcze chętnie obejrzeli i czego słuchali – i to samo zaproponować kolejnym.

Spotify i Netflix nie zamierzają bynajmniej podsunąć ci filmu czy utworu idealnego; doskonałość mało ich obchodzi. Tygodniowa playlista nowych utworów Spotify nie obiecuje, że wynajdzie ten jeden jedyny zespół na świecie, który będzie bez reszty pasował do twoich gustów i twojego nastroju. Algorytmy rekomendujące podsuwają ci jedynie te filmy i piosenki, które raczej nie powinny cię rozczarować – to oferty bezpiecznego sposobu spędzenia czasu wolnego. Zdarzy się, że algorytm podsunie ci coś, w czym się po prostu zakochasz, ale to ciągle bardziej wróżenie z fusów niż typowanie pewników. Tobie wystarczy, że poczujesz czasem euforię związaną z odkryciem dla siebie nowej muzyki. Algorytmy nie muszą mieć bezustannie najlepszych propozycji.

Podobieństwo doskonale się sprawdza w algorytmach rekomendacji. Kiedy jednak poprosisz algorytm o stworzenie dzieła sztuki bez podawania czystej miary jakości, zacznie się robić ciekawie. Czy komputer może być twórczy, jeżeli całe jego wyobrażenie o sztuce opiera się na tym, co już kiedyś było?

Dobrzy artyści pożyczają, wielcy artyści kradną – Pablo Picasso

W październiku 1997 roku publiczność zgromadzona w University of Oregon miała wysłuchać niecodziennego koncertu. Do stojącego na scenie samotnego fortepianu podeszła pianistka Winifred Kerner i usiadła przy instrumencie, żeby zagrać trzy krótkie utwory.

Pierwszym była mniej znana kompozycja na fortepian autorstwa mistrza baroku Johanna Sebastiana Bacha, drugą napisał w stylu bachowskim Steve Larson, profesor muzyki na uczelni goszczącej koncert,

trzecia zaś była autorstwa algorytmu zaprogramowanego do imitacji stylu Bacha.

Po wysłuchaniu wszystkich trzech kompozycji poproszono słuchaczy o przypisanie utworów do kompozytorów. Steve Larson był zdruzgotany – większa część publiczności uznała, że to właśnie jego utwór był dziełem komputera. Z kolei następny komunikat widownia przyjęła z mieszaniną przerażenia i rozbawienia: utwór, który słuchacze uznali za oryginalnego Bacha, był zwykłą imitacją napisaną przez maszynę.

W wywiadzie, którego niedługo po eksperymencie udzielił „New York Timesowi", Larson powiedział: „Mój podziw dla muzyki [Bacha] jest głęboki i kosmiczny. Ludzie dali się oszukać jakiemuś programowi na komputer i to było po prostu żenujące".

Nie tylko Larson był zbulwersowany. David Cope, który stworzył ów niezwykły algorytm do komponowania muzyki, widział już wcześniej podobne reakcje. „[Najpierw] bawiłem się w tę «grę» z wybranymi osobami – powiedział mi. – Kiedy udzielały błędnych odpowiedzi, wpadały w złość. Wściekały się na mnie już za sam pomysł. Bo ludzie uważają, że kreatywny może być tylko człowiek"[19].

Z tą opinią na pewno zgadzał się kognitywista i pisarz Douglas Hofstadter, który był organizatorem koncertu. W napisanej w 1979 roku książce *Gödel, Escher, Bach*, za którą dostał Nagrodę Pulitzera, wyraził się jasno: „Muzyka to język uczuć i dopóki programy nie będą miały tak złożonych uczuć jak my, nie ma mowy, żeby były zdolne skomponować coś równie pięknego (...). Sama myśl, że moglibyśmy zaprogramowanej «katarynce» kazać produkować rzeczy porównywalne z kompozycjami Bacha, to groteskowe i haniebne niezrozumienie głębi ludzkiego ducha"[20].

Ale po wysłuchaniu dzieł algorytmu Cope'a – znanych jako eksperymenty z inteligencji muzycznej (ang. *Experiments in Musical Intelligence*, EMI) – Hofstadter przyznał, że może rzeczywiście sprawa nie jest aż tak prosta: „Po EMI jestem skonsternowany i zaniepokojony – stwierdził. – Na tym etapie mogę się pocieszać wyłącznie świadomością, że EMI nie tworzy własnego stylu. Potrafi jedynie imitować innych kompozytorów. Ale mimo wszystko niewielka to pociecha. Jestem

absolutnie zdruzgotany, bo wygląda na to, że muzyka [być może] reprezentuje sobą znacznie mniej, niż do tej pory sądziłem”[21].

A więc jak to jest? Czy estetyczna wybitność pozostaje jedynym wyznacznikiem ludzkiej domeny? Czy może algorytm potrafi tworzyć sztukę? Skoro publiczność nie była zdolna odróżnić kompozycji EMI od muzyki wielkiego mistrza baroku, to czy można powiedzieć, że komputer błysnął prawdziwą kreatywnością?

Spróbujmy zmierzyć się z tymi pytaniami po kolei – rozpoczniemy od ostatniego. Jeśli chcemy na nie odpowiedzieć mądrze, warto poświęcić chwilę na zrozumienie funkcjonowania muzycznego algorytmu[25*]. David Cope dokładnie mi to wytłumaczył.

Pierwszym krokiem było przełożenie muzyki Bacha na elementy, które maszyna jest w stanie zrozumieć: „Musisz umieścić w bazie danych pięć wymiarów pojedynczej nuty: umiejscowienie w metrum, trwanie, wysokość, głośność i instrument”. Każdą nutę z katalogu dzieł zebranych Bacha Cope musiał właśnie tak, w postaci pięciu liczb, mozolnie wprowadzić do komputera. Samych tylko chorałów było 371, a to oznacza wiele harmonii, dziesiątki tysięcy nut, pięć liczb na każdą z nich. Była to benedyktyńska praca: „Całymi miesiącami nie robiłem nic poza wklepywaniem tych liczb do komputera. Ale taki już jestem, obsesje to moja specjalność”.

Kolejnym krokiem było przeanalizowanie następstwa każdego dźwięku. Dla każdej nuty w dowolnym chorale Bacha Cope sporządził wykaz wszystkich następników. Wszystkie te dane były zapisywane w swoistym repozytorium – banku dźwięków, w którym algorytm mógł sprawdzić każdy akord i znaleźć wyczerpującą listę wszystkich możliwych miejsc, do których wiodło dalej muzykę pióro Jana Sebastiana Bacha.

W tym sensie EMI zdradza pewne podobieństwa do algorytmów przewidujących, które znajdziemy w klawiaturze swojego smartfona. W oparciu o wcześniej zapisane zdania telefon tworzy listę słów, które prawdopodobnie będziemy chcieli wybrać w wiadomości, i wyświetla je jako sugestie w pasku nad klawiaturą[26*].

W ostatnim kroku trzeba było popuścić maszynie cugli. Cope ładował do systemu początkowy akord i instruował algorytm, żeby ze swojego

repozytorium losowo wybierał kolejne akordy. Potem proces się powtarzał – każdy następny akord był wybierany losowo z listy możliwych kontynuacji. W rezultacie powstała oryginalna kompozycja, która brzmi zupełnie tak, jakby napisał ją sam Bach[27*].

A może to j e s t Bach? Tak przynajmniej uważa Cope. „Wszystkie akordy są dziełem Bacha. To tak jakby wziąć kawałek parmezanu i zetrzeć go na tarce, a potem sklejać te płatki z powrotem w jedną całość. Cokolwiek z tego wyjdzie, będzie to zawsze parmezan".

Bez względu na to, komu ostatecznie przypisać autorstwo takich kompozycji, jedna rzecz jest pewna. Jakkolwiek pięknie mogą brzmieć utwory EMI, są to jedynie wyniki rekombinacji już istniejących dzieł. Komputer bardziej imituje prawidłowości, które odkrywa w muzyce Bacha, niż sam komponuje oryginalną muzykę.

Całkiem niedawno powstały nowe algorytmy tworzące miłą dla ucha muzykę, która jest już czymś więcej niż tylko zwykłą rekombinacją istniejących utworów. Szczególnie dobrze radziły sobie algorytmy genetyczne, swoiste maszyny uczące się, które wykorzystują prawidłowości doboru naturalnego. (Wystarczy spojrzeć na takiego pawia, żeby się przekonać, że ewolucja to i owo wie o tworzeniu piękna).

Idea jest prosta: algorytmy genetyczne traktują nuty jak DNA muzyki. Wszystko zaczyna się od wstępnej populacji „piosenek" – wybranych na chybił trafił i połączonych ze sobą nut. Na przestrzeni wielu „pokoleń" kolejnych utworów algorytm odnajduje i premiuje cechy „piękna" w muzyce, żeby z czasem generować coraz „lepsze" kompozycje. „Piękno" i „lepsze" ujmuję w cudzysłów, bo jak już wiem, nie można obiektywnie stwierdzić, co te słowa znaczą. Algorytmy genetyczne mogą tworzyć nie tylko muzykę, lecz także wiersze i obrazy, zawsze jednak będą bazować wyłącznie na mierze podobieństwa do tego, co już zostało stworzone.

I czasem niczego więcej nie trzeba. Jeżeli szukasz dla swojej strony internetowej czy filmiku na YouTubie podkładu muzycznego, który ma brzmieć jak piosenka ludowa, mało cię obchodzi, że będzie przypominać najlepsze istniejące piosenki tego typu; nie chce ci się komponować niczego nowego, ale nie chcesz też płacić tantiem autorom już istniejących utworów. Jeśli szukasz czegoś takiego, jest kilka firm, które ci w tym pomogą. Brytyjskie start-upy Jukebox i AI Music już oferują takie usługi,

korzystając z algorytmów komponujących muzykę; na pewno znajdziesz tam coś dla siebie. Niektóre utwory będą brzmieć jakby całkiem oryginalnie (no, prawie). Niektóre mogą być nawet piękne. Algorytmy może nie są wielkimi innowatorami, ale potrafią za to wspaniale naśladować.

Nie jest to bynajmniej opinia krzywdząca. Większość muzyki komponowanej przez nas również nie jest szczególnie innowacyjna. Wspominany już Armand Leroi, biolog ewolucyjny, który studiował kulturalną ewolucję muzyki pop, twierdzi, że mamy trochę zbyt romantyczną wizję zdolności kreatywnych człowieka. Nawet największe hity list przebojów mogą być dziełem maszyn. Oto co Leroi ma do powiedzenia o piosence *Happy* Pharrella Williamsa (coś mi się zdaje, że nie jest jego fanem): „*Happy, happy, happy, I'm so happy*... Proszę cię! To coś ma raptem pięć słów na krzyż! Trudno sobie wyobrazić utwór bardziej mechaniczny, który w jeszcze bardziej prymitywny sposób zaspokaja podstawowe ludzkie pragnienie znalezienia sobie na lato beztroskiej przyśpiewki na poprawę nastroju. Najgłupsza i najbardziej ograniczona piosenka świata. Jeśli więc to jest nasz poziom – nie tak trudno go osiągnąć".

Leroi nie ma też dobrego zdania o pisarskich umiejętnościach Adele: „Gdyby chciało ci się przeanalizować jej dowolny utwór, nie znalazłabyś w nim niczego, z czym by sobie nie poradził generator smutnych tekstów".

Możesz się z nim nie zgodzić (nie jestem pewna, czy sama bym się zgodziła), ale z pewnością jest coś na rzeczy, że ludzka kreatywność – podobnie jak wytwory algorytmów „komponujących" – to w zasadzie nowatorska kombinacja cudzych pomysłów. Jak mówi Mark Twain: „Nie ma czegoś takiego jak nowa idea. To rzecz niemożliwa. Bierzemy po prostu garść starych pomysłów i wrzucamy je do naszego mentalnego kalejdoskopu. Kiedy nim obracamy, powstają w nim nowe, ciekawe kombinacje. Obracamy tak więc i obracamy, tworząc bez przerwy nowe kombinacje; są to jednak wciąż te same kolorowe szkiełka, którymi ludzie się bawią od czasów najdawniejszych"[22].

Tymczasem Cope ma bardzo prostą definicję kreatywności, która bez trudu obejmuje także algorytmy: „Kreatywność to po prostu umiejętność

odnajdywania powiązań między dwiema rzeczami, które normalnie wydają się niepowiązane ze sobą".

Być może. Ciągle jednak nie mogę opędzić się od myśli, że jeśli EMI i inne podobne algorytmy wykazują się kreatywnością, stoi ona raczej na niskim poziomie. Ich muzyka może być piękna, ale nie jest głęboka. I choć bardzo bym się starała, nie potrafiłabym przestać myśleć, że uznając wytwory tych maszyn za sztukę, fundujemy sobie wersję świata o znacznie zubożonej zawartości kultury. Kulturową papkę dla zabicia stresu, a nie sztukę przez duże S.

Zbierając materiały do tej książki, uzmysłowiłam sobie, że moje wątpliwości związane z algorytmami tworzącymi sztukę wynikają z czegoś innego. Rzeczywiste pytanie nie brzmi: czy maszyny mogą być kreatywne? Bo mogą. Musimy za to spytać, co w ogóle uznajemy za sztukę.

Jestem matematyczką. Potrafię żonglować faktami o wynikach fałszywie pozytywnych i bez mrugnięcia powieką wygłaszać prawdy absolutne o ścisłości w statystyce. Ale w sferze sztuki byłabym skłonna zdać się na opinię Lwa Tołstoja. Zgadzam się z nim, że sednem prawdziwej sztuki jest połączenie serc, komunikowanie uczuć: „Jeżeli widzowie, słuchacze zarażają się takim samym uczuciem, jakie przeżywał twórca, jest to właśnie sztuka"[23]. Jeśli się zgodzisz z tezą Tołstoja, to mamy powód, dla którego maszyny nie są w stanie tworzyć prawdziwej sztuki. Powód pięknie wyrażony przez Douglasa Hofstadtera wiele lat przed jego spotkaniem z EMI: „«Program», który potrafiłby tworzyć muzykę (...) musiałby wędrować samodzielnie przez świat, z trudem pokonując labirynt życia i wyraźnie odczuwając każdą jego chwilę. Musiałby rozumieć radość i samotność zimnego nocnego wiatru, tęsknotę za ukochaną dłonią, nieosiągalność odległego miasta, zdruzgotanie i odnowę po śmierci bliskiego człowieka. Musiałby znać smak rezygnacji i znużenia życiem, smutku i rozpaczy, determinacji i zwycięstwa, bogobojności i nabożnej czci. Musiałby umieć pogodzić w sobie takie skrajności, jak nadzieja i lęk, cierpienie i wesołość, błogość i napięcie. To wszystko nie mogłoby się obyć bez poczucia wdzięku, humoru, rytmu, bez doświadczenia nieoczekiwanego – i rzecz jasna, bez owej czarodziejskiej, przemożnej świadomości tworzenia czegoś nowego. Tutaj – i tylko tutaj – można szukać źródeł znaczenia w muzyce"[24].

Mogę się jednak mylić. Być może kiedy sztuka algorytmiczna nabierze pozorów prawdziwego wytworu człowieka – jak EMI – zaczniemy ją cenić i dostrzegać w niej jakiś sens. Długa historia fabrykowanego popu zdaje się dowodzić, że ludzie mogą przejawiać reakcje emocjonalne, które z pozoru tylko wskazują na zespolenie twórcy z odbiorcą. Kto wie, czy kiedy te algorytmiczne dzieła sztuki bardziej się upowszechnią, a my staniemy się bardziej świadomi tego, że nie są dziełem człowieka, ta jednokierunkowość przeżycia przestanie nam przeszkadzać. Ostatecznie ludzi często łączą więzi emocjonalne z obiektami, które nie odwzajemniają miłości – ukochanymi pluszakami z dzieciństwa czy pająkami z domowego terrarium.

Dla mnie jednak prawdziwa sztuka nie może być dziełem przypadku. Są granice, których algorytmy nigdy nie przekroczą. Rzeczy, których nie da się skwantyfikować. Dane i statystyki potrafią mi powiedzieć rzeczy zdumiewające, o których się nie śniło filozofom, lecz nigdy nie powiedzą mi tego, jak to jest być człowiekiem.

Zakończenie

Rahinah Ibrahim była architektem, miała czworo dzieci, męża za oceanem, wolontariat w miejscowym szpitalu i nieukończony doktorat na Uniwersytecie Stanforda. Jakby tego było mało, ostatnio musiała się w trybie pilnym poddać histerektomii i choć w zasadzie doszła już do siebie, ciągle trudno jej było ustać o własnych siłach bez leków. Mimo to kiedy ogłoszono termin dorocznej, 38. Międzynarodowej Konferencji Systemologicznej, kobieta zarezerwowała lot na Hawaje, żeby w styczniu 2005 roku zaprezentować swój najnowszy artykuł na szerszym forum naukowym[1].

Drugiego stycznia 2005 roku, kiedy Ibrahim przybyła z córką na lotnisko w San Francisco, podeszła do okienka, podała urzędnikowi swoje dokumenty i poprosiła o zorganizowanie dla siebie wózka inwalidzkiego. Jej prośba nie została spełniona. Nazwisko kobiety wyświetliło się na ekranie komputera na federalnej liście osób z zakazem podróżowania; była to baza danych, w której po zamachu w 2001 roku na nowojorskie wieże umieszczono personalia osób podejrzewanych o terroryzm.

Przerażona córka Ibrahim, którą pozostawiono samą przy okienku, zadzwoniła do przyjaciela rodziny i powiadomiła go, że matkę zakuto w kajdanki i gdzieś zabrano. Ibrahim została zaprowadzona do radiowozu, posadzona na tylnym siedzeniu i zawieziona na posterunek policji. Zrewidowano ją, zaglądając pod hidżab, odmówiono jej leków i zamknięto ją w celi. Dwie godziny później na posterunek przybył agent z Departamentu Bezpieczeństwa Wewnętrznego z nakazem zwolnienia kobiety z aresztu i informacją, że jej nazwisko zostało usunięte z listy federalnej. Ibrahim zdążyła na konferencję na Hawajach, a potem poleciała do ojczystej Malezji odwiedzić rodzinę.

Kobieta znalazła się na liście, bo agent FBI zakreślił niewłaściwą kratkę w formularzu. Prawdopodobnie pomylił ze sobą dwie całkiem różne organizacje – Jemaah Islamiyah, ugrupowanie terrorystyczne, które

„wsławiło się" atakami bombowymi na Bali w 2002 roku, i Jemaah Islam, malezyjskie stowarzyszenie zawodowe dla osób studiujących za granicą, do którego należała Ibrahim; kobieta nie miała nigdy nic wspólnego z terroryzmem. Prosty błąd, który miał jednak dramatyczne konsekwencje. Bo gdy tylko trafił do systemu komputerowego, nabrał aury niepodważalnej prawdy i całkowicie zablokował wszelkie próby jego sprostowania. Historia bynajmniej nie skończyła się w terminalu lotniska w San Francisco.

Dwa miesiące później Ibrahim wracała z Malezji do domu i ponownie została zatrzymana na lotnisku. Tym razem sprawa nie wyjaśniła się tak szybko. Z powodu podejrzeń o powiązania z terroryzmem kobiecie cofnięto wizę. Mimo że była matką obywatelki amerykańskiej, miała dom w San Francisco i była zatrudniona na jednym z najbardziej prestiżowych uniwersytetów w kraju, nie pozwolono jej wrócić do Stanów Zjednoczonych. Musiało minąć prawie dziesięć lat, zanim Ibrahim ostatecznie wygrała sprawę w sądzie i oczyściła swoje imię. Prawie dziesięć lat, kiedy nie wolno jej było postawić stopy na amerykańskiej ziemi. A wszystko przez jeden ludzki błąd i wszechwładną maszynę.

Człowiek plus maszyna

Nikt nie wątpi w to, że automatyka wywarła głęboki i pozytywny wpływ na życie nas wszystkich. Algorytmy, które do tej pory skonstruowaliśmy, mogą się poszczycić oszałamiającymi wprost sukcesami. Pomagają nam diagnozować raka piersi, łapać seryjnych morderców i unikać katastrof lotniczych; dają każdemu z nas darmowy i łatwy dostęp do skarbnicy ludzkiej wiedzy – dzięki nim mamy ją na wyciągnięcie ręki; umożliwiają natychmiastową łączność między ludźmi na całym świecie w sposób niewyobrażalny dla naszych przodków. Wydaje się jednak, że w tym pędzie do automatyzacji, w tym pośpiechu, żeby rozwiązać jak najwięcej problemów naszego świata, wpadliśmy z deszczu pod rynnę. Algorytmy, choć są pożyteczne, a ich osiągnięcia imponujące, zostawiły nam do rozwiązania kilka grubych supłów.

Gdziekolwiek spojrzeć – wymiar sprawiedliwości, służba zdrowia, policja, nawet handel w internecie – borykamy się z naruszeniami

prywatności, z uprzedzeniami, błędami, brakiem odpowiedzialności i przejrzystości; niełatwo przewidzieć, kiedy to się skończy. Już samo pojawienie się algorytmów zachwiało poczuciem uczciwości, które dotyka istoty naszego człowieczeństwa i każe nam zadawać sobie pytania o to, jak ma wyglądać nasze społeczeństwo i jak skutecznie sobie radzić z nadciągającą erą wszechwładzy beznamiętnej techniki.

A może właśnie w tym problem? Może właśnie tutaj popełniamy błąd – uważając algorytmy za kolejną władzę?

Przede wszystkim nasza niechęć do kwestionowania możliwości algorytmu otworzyła furtkę ludziom, którzy są gotowi nas wykorzystywać. Zbierając materiały do tej książki, co rusz natrafiałam na wszelkiej maści hochsztaplerów i mitomanów żerujących na ludzkiej naiwności. Jedni wciskają policji i innym służbom algorytmy mające rzekomo „przewidywać", czy ktoś jest terrorystą albo pedofilem, wyłącznie na podstawie rysów twarzy – mimo mocnych dowodów naukowych, które obalają podobne mity. Drudzy bez mrugnięcia powieką twierdzą, że ich algorytm zasugeruje zmiany w jednym wierszu scenariusza, dzięki którym film przyniesie większe zyski[28*]. Jeszcze inni śmiało głoszą, że ich algorytm odnajdzie twoją jedyną prawdziwą miłość[29*].

Lecz nawet algorytmy, które faktycznie robią to, co twierdzą ich konstruktorzy, często nadużywają danej im władzy. W tej książce aż się roi od historii szkód, jakie mogą wyrządzić komputery. „Instrument budżetowy" wykorzystywany do arbitralnych cięć pomocy finansowej dla niedołężnych mieszkańców stanu Idaho. Algorytmy do obliczania ryzyka recydywy, które z przyczyn historycznych są skłonne uznawać za bardziej niebezpiecznych czarnoskórych podsądnych. System do wykrywania ostrego uszkodzenia nerek, który wszedł w posiadanie najbardziej wrażliwych i intymnych danych milionów pacjentów bez ich wiedzy ani zgody. Algorytm sieci supermarketów, przez który nastolatka straciła szansę powiadomienia ojca o tym, że zaszła w ciążę. Algorytm o nazwie Strategic Subject List, który miał pomagać ofiarom napadów z bronią w ręku, a był wykorzystywany przez policję jako lista podejrzanych. Przykłady niesprawiedliwego traktowania mamy po prostu wszędzie.

Wytykając jednak ciągle błędy maszyn, ryzykujemy sugestię, że istnieje jakaś idealna alternatywa, do której dążymy. Bardzo się starałam znaleźć

choćby jeden przykład algorytmu w pełni sprawiedliwego. Nawet te, które wyglądają z pozoru dobrze – jak autopilot samolotu czy sieci neuronowe diagnozujące raka – mają własne problemy na głębszych poziomach. Z rozdziału *Samochody* wiemy już, że autopilot może narazić osoby uczone prowadzenia zautomatyzowanych pojazdów na poważne niebezpieczeństwo w sytuacji realnego zagrożenia na drodze czy w powietrzu. Pojawiły się nawet głosy, że te rzekomo cudowne algorytmy odnajdywania guzów, które analizowaliśmy w rozdziale *Medycyna*, nie działają z równą skutecznością we wszystkich grupach etnicznych. Z drugiej strony, nie znajdziemy znów aż tylu przykładów doskonale uczciwych i sprawiedliwych systemów bez algorytmów. Gdziekolwiek spojrzeć, jakąkolwiek dziedzinę wziąć pod lupę, wystarczy sięgnąć wystarczająco głęboko, żeby natrafić na takie czy inne uprzedzenie.

A więc spróbujmy sobie wyobrazić, co się stanie, jeżeli przyjmiemy, że nie ma czegoś takiego jak system doskonały. Algorytmy b ę d ą robić błędy. Algorytmy b ę d ą nieuczciwe. To nie powinno nas w żadnej mierze zniechęcać do podejmowania wszelkich wysiłków, żeby stawały się coraz dokładniejsze i coraz mniej uprzedzone – być może jednak pogodzenie się z tym, iż algorytmy, zupełnie jak ludzie, nie są idealne, będzie miało ten skutek, że przestaniemy tak mocno wierzyć w ich wyroki.

Wyobraźmy sobie, że zamiast skupiać się wyłącznie na projektowaniu algorytmów, które będą się trzymać praktycznie nieosiągalnego standardu doskonałej uczciwości, zaczniemy projektować je tak, żeby łatwiej było wynagrodzić szkody spowodowane nieuniknionymi ich błędami; że zaczniemy również usilnie starać się o to, żeby kwestionowanie decyzji automatów było równie łatwe jak wcielanie ich w życie. Niewykluczone, że najlepszym rozwiązaniem będzie konstruowanie takich algorytmów, które będzie można sprawdzać na każdym etapie funkcjonowania. Wyobraźmy sobie, że będziemy je budować raczej po to, żeby wspierały ludzi w podejmowaniu decyzji, a nie po to, żeby nas instruowały, co powinniśmy robić. Żeby potrafiły pokazać nam ścieżkę prowadzącą do każdej swojej decyzji, a nie tylko informowały nas o wynikach obliczeń.

W moim odczuciu najlepsze algorytmy to takie, które uwzględniają człowieka na każdym etapie swojego działania. Takie, które wiedzą, że mamy zły nawyk traktowania maszyn jako nieomylne, i takie, które są

świadome własnych wad, więc z dumą prezentują nam swoją niedoskonałość.

To była jedna z najlepszych cech Watsona – maszyny skonstruowanej przez IBM do wygrywania w *Jeopardy!* Mimo że format teleturnieju wymagał od uczestników wybrania ostatecznie jednej odpowiedzi, algorytm prezentował także kilka możliwości, na które natrafił w procesie poszukiwania odpowiedzi, i podawał punktację odzwierciedlającą prawdopodobieństwo ich poprawności. Być może gdyby algorytmy obliczające ryzyko recydywy postępowały podobnie, sędziom byłoby łatwiej kwestionować generowane przez nie wyniki. I niewykluczone, że gdyby algorytmy rozpoznawania twarzy prezentowały kilka możliwych dopasowań, zamiast skupiać się na dobraniu idealnej pary, błędne identyfikacje nie byłyby tak wielkim problemem.

To właśnie ta cecha daje tak wielką skuteczność sieciom neuronowym, które analizują skrawki histopatologiczne pod kątem raka piersi. Algorytm nie stwierdza kategorycznie, które pacjentki mają guzy. Jego zadanie polega wyłącznie na wyłapaniu w gigantycznej masie komórek kilku podejrzanych skupisk, na które powinien zwrócić uwagę patolog. Algorytm nigdy się nie męczy, a patolog rzadko stawia niewłaściwą diagnozę. Maszyna i człowiek pracują razem, korzystając ze swoich mocnych stron i eliminując słabości.

Są też inne przykłady, także ze świata szachów, od czego zaczęła się ta książka. Garry Kasparow po przegranym meczu z Deep Blue nie odwrócił się od komputerów. Wręcz przeciwnie. Stał się wielkim orędownikiem idei „szachów centaura", w których rywalizują ze sobą dwa hybrydowe zespoły złożone z człowieka i komputera. Algorytm ocenia możliwe konsekwencje każdego ruchu, minimalizując ryzyko popełnienia głupiego błędu, grę jednak kontroluje człowiek.

„Kiedy gramy z pomocą komputera, możemy się skoncentrować na planowaniu strategii, nie tracąc tyle czasu na obliczenia – powiedział Kasparow. – Ludzka kreatywność w takich warunkach wybija się na pierwszy plan"[2]. Tak powstają szachy grane na poziomie wyższym niż kiedykolwiek wcześniej. Bezbłędna gra taktyczna i piękne, sensowne strategie. Najlepsze elementy obu światów.

Tak wygląda moja wymarzona przyszłość. Przyszłość bez aroganckich, tyrańskich algorytmów, których pełno na kartach tej książki. Przyszłość, w której przestaniemy postrzegać maszyny jako obiektywnych sędziów i zaczniemy odnosić się do nich tak, jak odnosilibyśmy się do każdej innej władzy – kwestionując jej decyzje, analizując jej motywy, rozumiejąc własne emocje, domagając się ujawnienia, kto zyska na jej postanowieniach, pociągając ją do odpowiedzialności i nie godząc się biernie na wszystkie jej zapędy. Wydaje mi się, że to jest klucz do przyszłości, w której ogólny wpływ algorytmów będzie stanowił pozytywną siłę w społeczeństwie. A żeby tak się stało, musimy sami zakasać rękawy. Nie wyobrażam sobie innego scenariusza, bo jedna rzecz jest pewna. Ludzie nigdy nie byli tak ważni jak teraz – w epoce algorytmów.

Podziękowania

Domyślam się, że są na świecie ludzie, którym pisanie przychodzi łatwo. Na pewno macie takich znajomych – wyskakują z łóżka przed wschodem słońca i do południa mają napisany rozdział; zapominają zejść na kolację, bo złapali taki ciąg twórczy, że czas przestał w ogóle istnieć.

To zdecydowanie nie ja.

Proces pisania był dla mnie codzienną walką z tą Hannah Fry, która po prostu chce rozsiąść się na kanapie przed telewizorem, otworzyć chipsy i włączyć Netfliksa; była to totalna wojna z tsunami paniki, z którą, jak sądziłam, pożegnałam się już definitywnie po napisaniu doktoratu. Tak naprawdę nie pisałam tej książki, tylko – kopiąc i wrzeszcząc – wyrywałam ją z siebie. Czasami dosłownie.

Tym bardziej jestem wdzięczna grupie wspaniałych ludzi, którzy chcieli mi w tym pomagać. Oto mój fantastyczny zespół wydawniczy, który nie szczędził mi czasu ani pomysłów w ciągu ostatniego roku: Susanna Wadeson, Quynh Do, Claire Conrad, Emma Parry, Gillian Somerscales, Emma Burton, Sophie Christopher, Hannah Bright, Caroline Saine; w tle pomagali mi ludzie z Janklow & Nesbit, Transworld i Norton. Podobnie Sue Rider, Kat Bee i Tom Copson. Bez Was bym się zgubiła.

Ogromne podziękowania kieruję też do moich informatorów. Kilkoro z nich wymieniłam w tekście, wszyscy jednak pomogli mi sformułować myśli zawarte w tej książce: Jonathan Rowson, Nigel Harvey, Adam Benforado, Giles Newell, Richard Berk, Sheena Urwin, Steyve Colgan, Mandeep Dhami, Adrian Weller, Toby Davies, Rob Jenkins, Jon Kanevsky, Timandra Harkness, Dan Popple i zespół z policji w West Midlands, Andy Beck, Jack Stilgoe, Caroline Rance, Paul Newman, Phyllis Illarmi, Armand Leoni, David Cope, Ed Finn, Kate Devlin, Shelia Hayman, Tom Chatwin, Carl Gombrich, Johnny Ryan, Jon Crowcroft i Frank Kelly.

Nie sposób pominąć Sue Webb i Debbie Enright z Network Typing oraz Sharon Richardson, Shruthi Rao i Willa Storra – ich pomoc w nagięciu tej

książki do właściwej formy była nieoceniona. I na koniec, kiedy zdołałam wreszcie sklecić ze słów coś w rodzaju zdań, to James Fulker, Elisabeth Adlington, Brendan Maginnis, Ian Hunter, Omar Miranda, Adam Dennett, Michael Veale, Jocelyn Bailey, Cat Black, Tracy Fry, Adam Rutherford i Thomas Oléron Evans pomogli mi wyłapać moje najgorsze maniery i odpowiednio je utemperować. A Geoff Dahl, który wspierał mnie moralnie przez cały ten czas, miał także świetny pomysł na okładkę.

Bardzo jestem wdzięczna moim redaktorom, oto oni: Elizabeth Cleverdon, Bethany Davies, Ben Dickson, Mike Downes, Charlie i Laura Galan, Katie Heath, Mia Kazi-Fornari, Fatah Ioualitene, Siobhan Mathers, Mabel Smaller, Ali Seyhun Saral, Jennifer Shelley, Edward Steele, Daniel Vesma, Jass Ubhi.

Jestem także niewyobrażalnie wdzięczna mojej rodzinie za przykładną lojalność i za to, że stała za mną murem. Phil, Tracy, Natalie, Marge & Parge, Omar, Mike i Tania – mieliście dla mnie więcej cierpliwości, niż na to zasługiwałam. (Choć nie bierzcie tego zbyt dosłownie, ponieważ prawdopodobnie napiszę kolejną książkę i znowu będę potrzebowała Waszej pomocy, OK?)

No i na samym końcu Ty, Edith – bynajmniej nie ostatnia. Szczerze mówiąc, w ogóle mi nie pomogłaś, ale tak po prostu było lepiej.

Podziękowania za udostępnienie zdjęć

s. 23 Pies-samochód, zdjęcie udostępnione za zgodą Danilo Vasconcellosa Vargasa, Kyushu University, Fukuoka, Japonia.

s. 119 Goryl w płucu, zdjęcie udostępnione za zgodą Traftona Drewa, University of Utah, Salt Lake City, Stany Zjednoczone.

s. 209 Zdjęcia Steve'a Talleya © Steve Talley (po lewej), FBI (z prawej strony).

s. 212 Neil Douglas i jego sobowtór, zdjęcie udostępnione za zgodą Neila Douglasa.

s. 218 Żółwiowe okulary, zdjęcie udostępnione za zgodą Mahmooda Sharifa, Carnegie Mellon University, Pittsburgh, Stany Zjednoczone; „Milla Jovovich na Festiwalu Filmowym w Cannes", Georges Biard.

Przypisy

Parę słów o tytule

[1] Brian W. Kernighan, Dennis M. Ritchie, *The C Programming Language*, Upper Saddle River, NJ 1978.

Wstęp

[1] Robert A. Caro, *The Power Broker: Robert Moses and the Fall of New York*, London 2015, s. 318.

[2] Zdecydowanie warto przeczytać dwa fantastyczne eseje poświęcone temu zagadnieniu. Pierwszy to: Langdon Winner, *Do artifacts have politics?*, „Daedalus" 1980, t. 109, nr 1, s. 121–136, https://www.jstor.org/stable/20024652 , w którym pojawia się przykład z mostami Mosesa. I bardziej współczesna wersja: Kate Crawford, *Can an algorithm be agonistic? Ten scenes from life in calculated publics*, „Science, Technology and Human Values" 2016, t. 41, nr 1, s. 77–92.

[3] „Scunthorpe Evening Telegraph", 9 kwietnia 1996.

[4] Chukwuemeka Afigbo (@nke_ise) zamieścił o tym filmik na Twitterze. Warto zobaczyć, jeśli jeszcze go nie widzieliście. Dostępny także na YouTube: https://www.youtube.com/watch?v=87QwWpzVy7I .

[5] Wywiad w CNN, Mark Zuckerberg, *I'm really sorry that this happened*, YouTube, 21 marca 2018, https://www.youtube.com/watch?v=G6DOhioBfyY .

Władza

[1] Z prywatnej rozmowy z arcymistrzem Jonathanem Rowsonem.

[2] Feng-Hsiung Hsu, *IBM's Deep Blue Chess grandmaster chips*, „IEEE Micro", 1999, t. 19, nr 2, s. 70–81, http://ieeexplore.ieee.org/document/755469/ .

[3] Garry Kasparov, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, London 2017.

[4] TheGoodKnight, *Deep Blue vs Garry Kasparov Game 2 (1997 Match)*, YouTube, 18 października 2012, https://www.youtube.com/watch?v=3Bd1Q2rOmok&t=2290s .

[5] Tamże.

[6] Steven Levy, *Big Blue's Hand of God*, „Newsweek", 18 maja 1997, http://www.newsweek.com/big-blues-hand-god-173076 .

[7] G. Kasparov, *Deep Thinking*, dz. cyt., s. 187.

[8] Tamże, s. 191.

[9] Według słownika Merriam-Webster. Definicja z *The Oxford English Dictionary* bardziej nawiązuje do matematycznej natury algorytmów: „proces lub zbiór reguł stosowany w obliczeniach lub innych procedurach rozwiązywania problemów, zwłaszcza przez komputer".

[10] Algorytmy można pogrupować na bardzo wiele sposobów i nie mam wątpliwości, że część informatyków będzie narzekać, że moja lista jest zbyt uproszczona. To prawda – bardziej wyczerpująca klasyfikacja musiałaby uwzględnić jeszcze mapowanie, redukcję, regresję i clustering. W końcu jednak zdecydowałam się na ten właśnie zestaw, zaczerpnięty z: Nicholas Diakopoulos, *Algorithmic Accountability Reporting: On the Investigation of Black Boxes* (New York 2014) – ponieważ świetnie wyjaśnia podstawy i skutecznie odczarowuje i ukonkretnia ten ogromny i bardzo złożony obszar badań.

[11] Kerbobotat, *Went to buy a baseball bat on Amazon, they have some interesting suggestions for accessories*, „Reddit", 28 września 2013, https://www.reddit.com/r/funny/comments/1nb16l/went_to_buy_a_baseball_bat_on_amazon_they_have/ .

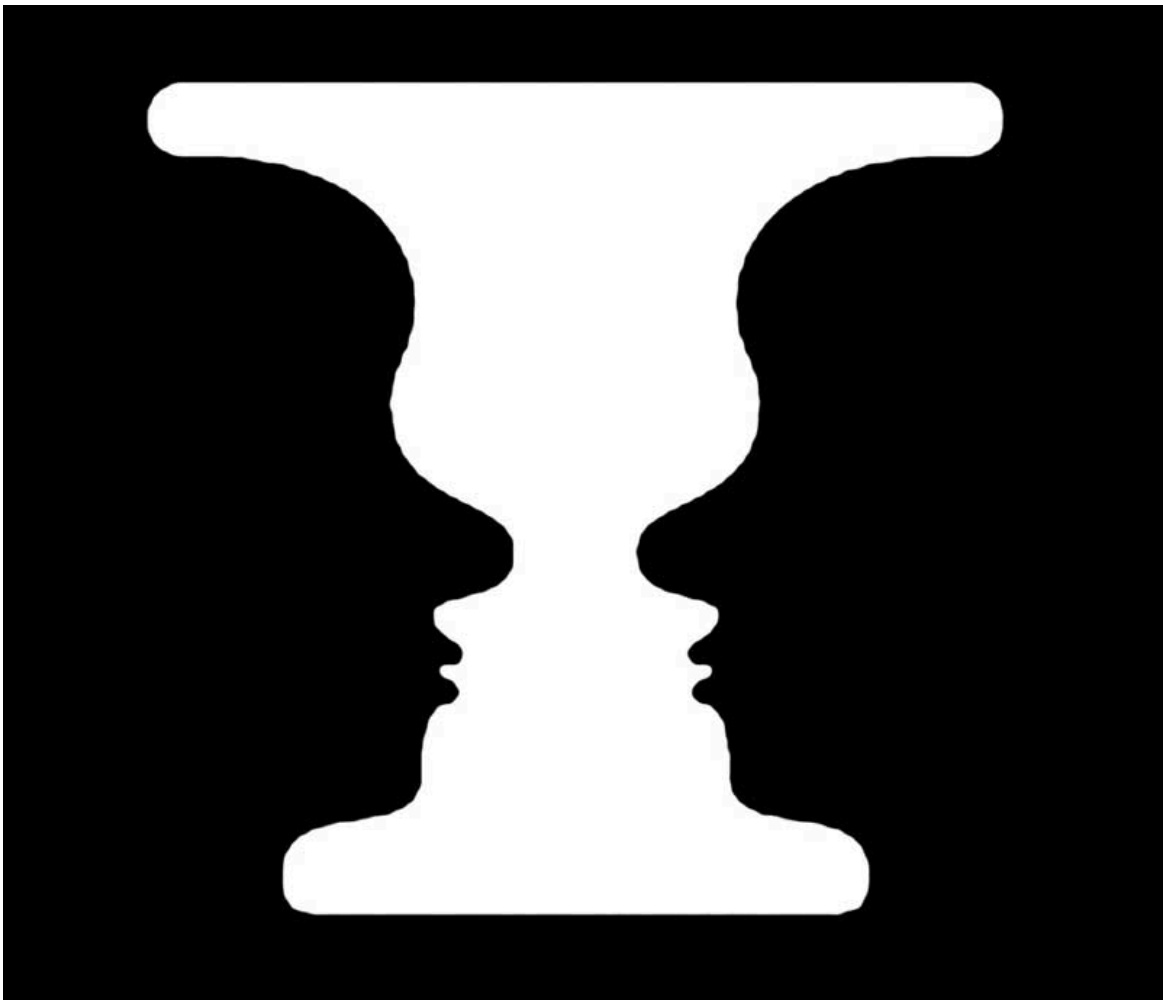
[12] Sarah Perez, *Uber debuts a „smarter" UberPOOL in Manhattan*, „TechCrunch", 22 maja 2017, https://techcrunch.com/2017/05/22/uber-debuts-a-smarter-uberpool-in-manhattan/ .

[13] Celowo użyłam sformułowania „teoretycznie", gdyż rzeczywistość może być nieco inna. Niektóre algorytmy powstawały przez wiele lat, pisane przez setki, a nawet tysiące informatyków, z których każdy miał swój udział w rozwoju kodu. Wraz z przybywaniem jego wierszy rośnie również złożoność systemu i wszystkie wątki logiczne plączą się niczym spaghetti na talerzu. Skutek jest taki, że śledzenie logiki algorytmu staje się niemożliwe – jest on tak skomplikowany, że żaden człowiek nie byłby w stanie go zrozumieć.

W 2013 roku firmie Toyota zasądzono karę 3 milionów dolarów w ramach odszkodowania za śmiertelny wypadek spowodowany przez jeden z wyprodukowanych przez nią samochodów; auto zaczęło przyspieszać w sposób niekontrolowany, mimo że kierująca pojazdem kobieta wciskała pedał hamulca, a nie gazu. Powołany na świadka ekspert orzekł przed sądem, że winą za wypadek należy obarczyć niezamierzoną instrukcję ukrytą w głębokich splotach programów komputerowych. Zob. Phil Koopman, *A case study of Toyota unintended acceleration and software safety*, Carnegie Mellon University, Pittsburgh, 18 września 2014, https://users.ece.cmu.edu/~koopman/pubs/koopman14_toyota_ua_slides.pdf .

[14] Złudzenie to (zamieszczony poniżej obrazek pochodzi z https://commons.wikimedia.org/wiki/File:Vase_of_rubin.png) znane jest jako „kielich Rubina" – od Edgara Rubina, który jest autorem pomysłu. To przykład obrazka w i e l o z n a c z n e g o – może przedstawiać dwie widoczne z profilu zwrócone do siebie czarne twarze albo jeden biały kielich. Patrząc na ten rysunek, stosunkowo łatwo przełączać widzenie z jednej możliwości na drugą, wystarczyłoby jednak kilka dodatkowych

kresek, żeby przechylić szalę interpretacji na jedną ze stron – zarys oczu na twarzach albo cień na obłej podstawie kielicha.



[15] Przykład z rozpoznawaniem na obrazku psa-samochodu to bardzo podobna historia. Zespół japońskich badaczy znalazł obrazek, który lokował się dokładnie na granicy dwóch klasyfikacji, oraz najmniejsze możliwe zakłócenie, które powodowało w interpretacji maszyny zmianę kategorii obrazka. Zob. Jiawei Su, Danilo Vasconcellos Vargas, Kouichi Sakurai, *One pixel attack for fooling deep neural networks*, arXiv:1719.08864v4 [cs.LG], 22 lutego 2018, https://arxiv.org/pdf/1710.08864.pdf .

[16] Chris Brooke, *I was only following satnav orders is no defence: driver who ended up teetering on cliff edge convicted of careless driving*, „Daily Mail", 16 września 2009, http://www.dailymail.co.uk/news/article-1213891/Driver-

ended-teetering-cliff-edge-guilty-blindly-following-sat-nav-directions.html#ixzz59vihbQ2n .

[17] Tamże.

[18] Robert Epstein, Ronald E. Robertson, *The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections*, „Proceedings of the National Academy of Sciences", 2015, t. 112, nr 33, s. E4512–E4521, http://www.pnas.org/content/112/33/E4512 .

[19] David Shultz, *Internet search engines may be influencing elections*, „Science", 7 sierpnia 2015, http://www.sciencemag.org/news/2015/08/internet-search-engines-may-be-influencing-elections .

[20] Robert Epstein, Ronald E. Robertson, *The search engine manipulation effect (SEME)*, dz. cyt.

[21] Linda J. Skitka, Kathleen Mosier, Mark D. Burdick, *Accountability and automation bias*, „International Journal of Human-Computer Studies" 2000, t. 52, s. 701–717, http://lskitka.people.uic.edu/IJHCS2000.pdf .

[22] KW v. Armstrong, US District Court, D. Idaho, 2 maja 2012, https://scholar.google.co.uk/scholar_case?case=17062168494596747089&hl=en&as_sdt=2006 .

[23] Jay Stanley, *Pitfalls of Artificial Intelligence Decision making Highlighted in Idaho ACLU Case*, ACLU, 2 lipca 2017, https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case .

[24] *K.W. v. Armstrong*, „Leagle.com", 24 marca 2014, https://www.leagle.com/decision/infdco20140326c20 .

[25] Tamże.

[26] ACLU of Idaho, Staff, https://www.acluidaho.org/en/about/staff .

[27] Jay Stanley, *Pitfalls of Artificial Intelligence Decision Making*, dz. cyt.

[28] ACLU, *Ruling mandates important protections for due process rights of Idahoans with developmental disabilities*, 30 marca 2016, https://www.aclu.org/news/federal-court-rules-against-idaho-department-health-and-welfare-medicaid-class-action .

[29] Jay Stanley, *Pitfalls of Artificial Intelligence Decision Making*, dz. cyt.

[30] Tamże.

[31] Tamże.

[32] Tamże.

[33] Tamże.

[34] Kristine Phillips, *The former Soviet officer who trusted his gut – and averted a global nuclear catastrophe*, „Washington Post", 18 września 2017, https://www.washingtonpost.com/news/retropolis/wp/2017/09/18/the-former-soviet-officer-who-trusted-his-gut-and-averted-a-global-nuclear-catastrophe/?utm_term=.6546e0f06cce .

[35] Pavel Aksenov, *Stanislav Petrov: the man who may have saved the world*, BBC News, 26 września 2013, http://www.bbc.co.uk/news/world-europe-24280831 .

[36] Tamże.

[37] Stephen Flanagan, *Re: Accident at Smiler Rollercoaster, Alton Towers, 2 June 2015: Expert's Report*, przygotowany na prośbę Komisji Zdrowia i Bezpieczeństwa, październik 2015, http://www.chiark.greenend.org.uk/~ijackson/2016/Expert_witness_report_from_Steven_Flanagan.pdf .

[38] Paul E. Meehl, *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Minneapolis 1996; pierwsze wydanie: 1954,

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.693.6031&rep=rep1&type=pdf .

[39] William M. Grove, David H. Zald, Boyd S. Lebow, Beth E. Snitz, Chad Nelson, *Clinical versus mechanical prediction: a meta-analysis*, „Psychological Assessment" 2000, t. 12, nr 1, s. 19.

[40] Berkeley J. Dietvorst, Joseph P. Simmons, Cade Massey, *Algorithmic aversion: people erroneously avoid algorithms after seeing them err*, „Journal of Experimental Psychology", wrzesień 2014, http://opim.wharton.upenn.edu/risk/library/WPAF201410-AlgorthimAversion-Dietvorst-Simmons-Massey.pdf .

Dane

[1] Nicholas Carlson, *Well, these new Zuckerberg IMs won't help Facebook's privacy problems*, „Business Insider", 13 maja 2010, http://www.businessinsider.com/well-these-new-zuckerberg-ims-wont-help-facebooks-privacy-problems-2010-5?IR=T .

[2] Clive Humby, Terry Hunt, Tim Phillips, *Scoring Points: How Tesco Continues to Win Customer Loyalty*, London 2008.

[3] Tamże, wersja na Kindle'a, 1313–1317.

[4] Zob. Eric Schmidt, *The creepy line*, YouTube, 11 lutego 2013, https://www.youtube.com/watch?v=0-rvER6YTss .

[5] Charles Duhigg, *How companies learn your secrets*, „New York Times", 16 lutego 2012, https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html .

[6] Tamże.

[7] Sarah Buhr, *Palantir has raised $880 million at a $20 billion valuation*, „TechCrunch", 23 grudnia 2015.

[8] Federal Trade Commission, *Data Brokers: A Call for Transparency and Accountability*, Washington, DC, maj 2014, https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf .

[9] Tamże.

[10] Wolfie Christl, *Corporate surveillance in everyday life*, Cracked Labs, lipiec 2017, http://crackedlabs.org/en/corporate-surveillance .

[11] Heidi Waterhouse, *The death of data: retention, rot, and risk*, The Lead Developer, Austin, Texas, 2 marca 2018, https://www.youtube.com/watch?v=mXvPChE09iU .

[12] Amit Datta, Michael Carl Tschantz, Anupam Datta, *Automated experiments on ad privacy settings*, „Proceedings on Privacy Enhancing Technologies" 2015, nr 1, s. 92–112.

[13] Latanya Sweeney, *Discrimination in online ad delivery*, „Queue" 2013, t. 11, nr 3, s. 10, https://dl.acm.org/citation.cfm?id=2460278 .

[14] Jon Brodkin, *Senate votes to let ISPs sell your Web browsing history to advertisers*, „Ars Technica", 23 marca 2017, https://arstechnica.com/tech-policy/2017/03/senate-votes-to-let-isps-sell-your-web-browsing-history-to-advertisers/ .

[15] Svea Eckert, Andreas Dewes, *Dark data*, DEFCON Conference 25, 20 października 2017, https://www.youtube.com/watch?v=1nvYGi7-Lxo .

[16] Tę część pracy badacze oparli na: Arvind Narayanan, Vitaly Shmatikov, *Robust de-anonymization of large sparse datasets*, prezentacja pokazana na IEEE Symposium on Security and Privacy, 18–22 maja 2008.

[17] Michal Kosinski, David Stillwell, Thore Graepel, *Private traits and attributes are predictable from digital records of human behavior*, „PNAS" 2013, t. 110, nr 15, s. 5802–5805.

[18] Tamże.

[19] Wu Youyou, Michal Kosinski, David Stillwell, *Computer-based personality judgments are more accurate than those made by humans*, „Proceedings of the National Academy of Sciences" 2015, t. 112, nr 4, s. 1036–1040.

[20] S.C. Matz, M. Kosinski, G. Nave, D.J. Stillwell, *Psychological targeting as an effective approach to digital mass persuasion*, „Proceedings of the National Academy of Sciences" 2017, t. 114, nr 48, 201710966.

[21] Paul Lewis, Paul Hilder, *Leaked: Cambridge Analytica's blueprint for Trump victory*, „Guardian", 23 marca 2018.

[22] *Cambridge Analytica planted fake news*, BBC, 20 marca 2018, http://www.bbc.co.uk/news/av/world-43472347/cambridge-analytica-planted-fake-news .

[23] Adam D.I. Kramer, Jamie E. Guillory, Jeffrey T. Hancock, *Experimental evidence of massive-scale emotional contagion through social networks*, „Proceedings of the National Academy of Sciences" 2014, t. 111, nr 24, s. 8788–8790.

[24] Jamie Bartlett, *Big data is watching you – and it wants your vote*, „Spectator", 24 marca 2018.

[25] Li Xiaoxiao, *Ant financial subsidiary starts offering individual credit scores*, „Caixin", 2 marca 2015, https://www.caixinglobal.com/2015-03-02/101012655.html .

[26] Rick Falkvinge, *In China, your credit score is now affected by your political opinions – and your friends' political opinions*, „Privacy News Online", 3 października 2015, https://www.privateinternetaccess.com/blog/2015/10/in-china-your-credit-score-is-now-affected-by-your-political-opinions-and-your-friends-political-opinions/ .

[27] *State Council Guiding Opinions Concerning Establishing and Perfecting Incentives for Promise-Keeping and Joint Punishment Systems for Trust-Breaking, and Accelerating the Construction of Social Sincerity*, China Copyright and Media, 30 maja 2016, uzupełnione 18 października 2016, https://chinacopyrightandmedia.wordpress.com/2016/05/30/state-council-guiding-opinions-concerning-establishing-and-perfecting-incentives-for-promise-keeping-and-joint-punishment-systems-for-trust-breaking-and-accelerating-the-construction-of-social-sincer/ .

[28] Rachel Botsman, *Who Can You Trust? How Technology Brought Us Together – and Why It Could Drive Us Apart*, London 2017, wersja na Kindle'a, s. 151.

Wymiar sprawiedliwości

[1] John-Paul Ford Rojas, *London riots: Lidl water thief jailed for six months*, „Telegraph", 7 stycznia 2018, http://www.telegraph.co.uk/news/uknews/crime/8695988/Londonriots-Lidl-water-thief-jailed-for-six-months.html .

[2] Matthew Taylor, *London riots: how a peaceful festival in Brixton turned into a looting free-for-all*, „Guardian", 8 sierpnia 2011, https://www.theguardian.com/uk/2011/aug/08/london-riots-festival-brixton-looting .

[3] John-Paul Ford Rojas, *London riots*, dz. cyt.

[4] Josh Halliday, *London riots: how BlackBerry Messenger played a key role*, „Guardian", 8 sierpnia 2011, https://www.theguardian.com/media/2011/aug/08/london-riotsfacebook-twitter-blackberry .

[5] David Mills, *Paul and Richard Johnson avoid prison over riots*, „News Shopper", 13 stycznia 2012, http://www.newsshopper.co.uk/londonriots/9471288.Father_and_son_avoid_prison_over_riots/ .

[6] Tamże.

[7] John-Paul Ford Rojas, *London riots*, dz. cyt. „Normalnie policja nie zatrzymałaby cię za takie wykroczenie. Nie wsadzono by cię do aresztu. Nie postawiono by cię przed sądem", napisała w 2015 roku Hannah Quirk, starszy wykładowca prawa karnego na uniwersytecie w Manchesterze (Carly Lightowlers, Hannah Quirk, *The 2011 English „riots": prosecutorial zeal and judicial abandon*, „British Journal of Criminology" 2015, t. 55, nr 1, s. 65–85).

[8] David Mills, *Paul and Richard Johnson avoid prison over riots*, dz. cyt.

[9] William Austin, Thomas A. Williams III, *A survey of judges' responses to simulated legal cases: research note on sentencing disparity*, „Journal of Criminal Law and Criminology" 1977, t. 68, nr 2, s. 306–310.

[10] Mandeep K. Dhami, Peter Ayton, *Bailing and jailing the fast and frugal way*, „Journal of Behavioral Decision Making" 2001, t. 14, nr 2, s. 141–168, http://onlinelibrary.wiley.com/doi/10.1002/bdm.371/abstract .

[11] W kwestii najlepszego rozwiązania w każdej ze spraw odmienne zdanie mogła mieć nawet połowa sędziów.

[12] Statystycy mają metodę określania tego rodzaju powtarzalności osądów – to tak zwany współczynnik Kappa Cohena. Bierze on pod uwagę to, że – nawet gdyby ktoś po prostu zgadywał – może się zdarzyć, że zostanie wydany osąd zgodny przez zwykły przypadek. Wynik równy jedności oznacza idealną powtarzalność. Wynik równy zeru oznacza, że równie dobrze moglibyśmy swoje orzeczenia losować. Wyniki sędziów w tym eksperymencie oscylowały między obiema skrajnościami, a średnia wyniosła 0,69.

[13] Diane Machin, *Sentencing guidelines around the world*, artykuł przygotowany na Scottish Sentencing Council, maj 2005, https://www.scottishsentencingcouncil.org.uk/media/1109/paper-31a-sentencing-guidelines-around-the-world.pdf .

[14] Tamże.

[15] Tamże.

[16] Ernest W. Burgess, *Factors determining success or failure on parole*, w: *The Workings of the Intermediate-Sentence Law and Parole System in Illinois*, Springfield, IL 1928. Niełatwo ten dokument namierzyć, polecam więc alternatywną lekturę autorstwa kolegi Burgessa, Tibbitsa, który kontynuował pierwotne badania (Clark Tibbitts, *Success or failure on parole can be predicted: a study of the records of 3,000 youths paroled from the Illinois State Reformatory*, „Journal of Criminal Law and Criminology" 1931, t. 22, nr 1, https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=2211&context=jclc). Pozostałe kategorie Burgessa to „czarna owca", „przestępca z przypadku", „bałwan" i „gangster". W jego klasyfikacji najmniejsze ryzyko recydywy stanowili „chłopcy ze wsi".

[17] Karl F. Schuessler, *Parole prediction: its history and status*, „Journal of Criminal Law and Criminology" 1955, t. 45, nr 4, s. 425–431, https://pdfs.semanticscholar.org/4cd2/31dd25321a0c14a9358a93ebccb6f15d3169.pdf .

[18] Tamże.

[19] Bernard E. Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*, Chicago–London 2007, s. 1.

[20] Philip Howard, Brian Francis, Keith Soothill, Les Humphreys, *OGRS 3: The Revised Offender Group Reconviction Scale*, Research Summary 7/09, London 2009, https://core.ac.uk/download/pdf/1556521.pdf .

[21] Słówko ostrzeżenia: bardzo możliwe, że statystyka ta jest jednak skażona błędem doboru próby. „Pytanie do publiczności" było zazwyczaj używane w pierwszych rundach gry, kiedy pytania były znacznie łatwiejsze niż później, gdy uczestnicy korzystali z „telefonu do przyjaciela". Niemniej jednak większa trafność opinii zbiorowych w stosunku do opinii jednostek to zjawisko dobrze udokumentowane. Więcej na ten temat zob. James

Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter than the Few*, New York 2004, s. 4.

[22] Netflix Technology Blog, https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5 .

[23] Shih-ho Cheng, *Unboxing the random forest classifier: the threshold distributions*, Airbnb Engineering and Data Science, https://medium.com/airbnb-engineering/unboxing-the-random-forest-classifier-the-threshold-distributions-22ea2bb58ea6 .

[24] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan, *Human Decisions and Machine Predictions*, Cambridge, MA 2017, NBER Working Paper nr 23180, http://www.nber.org/papers/w23180 . W badaniu tym wykorzystano faktycznie „drzewa decyzji z optymalizacją gradientową", algorytm podobny do lasu losowego. Obie metody w procesie decyzyjnym łączą przewidywania całej masy drzew, z tą różnicą, że w metodzie optymalizacji gradientowej drzewa „sadzi się" sekwencyjnie, podczas gdy w lasach losowych drzewa są rozmieszczone równolegle. Badacze najpierw podzielili bazę danych na dwie równe części. Jedną połówkę wykorzystano do uczenia algorytmu, a drugą odłożono tymczasem na bok. Kiedy algorytm był gotowy, zabrał się do oceny przypadków z drugiej połówki – której wcześniej „nie widział" – przewidując rozwój zdarzeń. (Bez uprzedniego podzielenia bazy danych algorytm byłby po prostu fantazyjną tablicą programistyczną).

[25] Badacze poświęcili sporo czasu na skonstruowanie technik statystycznych, które poradziłyby sobie z tym problemem – aby mimo wszystko dało się sensownie porównać prognozy sędziów i algorytmów. Więcej informacji na ten temat zob. tamże.

[26] *Costs per place and costs per prisoner by individual prison: National Offender Management Service Annual Report and Accounts 2015–16, Management Information Addendum*, Ministry of Justice Information Release, 27

października 2016, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/563326/costs-per-place-cost-per-prisoner-2015-16.pdf .

[27] Marc Santora, *City's annual cost per inmate is $168,000, study finds*, „New York Times", 23 sierpnia 2013, http://www.nytimes.com/2013/08/24/nyregion/citys-annual-cost-per-inmate-is-nearly-168000-study-says.html ; Harvard University, „Harvard at a glance", https://www.harvard.edu/about-harvard/harvard-glance .

[28] Luke Dormehl, *The Formula: How Algorithms Solve all our Problems... and Create More*, London 2014, s. 123.

[29] Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner, *Machine bias*, ProPublica, 23 maja 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing .

[30] Kwestionariusz „oceny ryzyka", https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html .

[31] Tim Brennan, William Dieterich, Beate Ehret (Northpointe Institute), *Evaluating the predictive validity of the COMPAS risk and needs assessment system*, „Criminal Justice and Behavior" 2009, t. 36, nr 1, s. 21–40, http://www.northpointeinc.com/files/publications/Criminal-Justice-Behavior-COMPAS.pdf . Jak wykazało badanie z 2018 roku, algorytm COMPAS-u ma podobną dokładność prognozowania jak „zespół" ludzi. Badacze zademonstrowali, że 20 niedoświadczonych osób, które poproszono o ocenę ryzyka recydywy, osiągnęło wynik porównywalny z wynikiem COMPAS-u. To interesujące porównanie, warto jednak pamiętać, że na prawdziwych procesach sędziowie nie korzystają z zespołu zupełnie obcych sobie ludzi, którzy na zapleczu głosują, co zrobić z oskarżonym. Sędziowie są na sali zupełnie sami. I jest to jedyne porównanie, które się naprawdę liczy. Zob. Julia Dressel, Hany Farid, *The accuracy, fairness, and limits of predicting recidivism*, „Science Advances", 2018 t. 4, nr 1.

[32] *Christopher Drew Brooks v. Commonwealth*, Court of Appeals of Virginia, Memorandum Opinion by Judge Rudolph Bumgardner III, 28 stycznia 2004, https://law.justia.com/cases/virginia/court-of-appeals-unpublished/2004/2540023.html .

[33] *ACLU brief challenges constitutionality of Virginia's sex offender risk assessment guidelines*, ACLU of Virginia, 28 października 2003, https://acluva.org/en/press-releases/aclu-brief-challenges-constitutionality-virginias-sex-offender-risk-assessment .

[34] *State v. Loomis*, Supreme Court of Wisconsin, 13 lipca 2016, http://caselaw.findlaw.com/wi-supreme-court/1742124.html .

[35] Cytowane wypowiedzi Richarda Berka pochodzą z moich rozmów i korespondencji z nim.

[36] Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner, *Machine bias*, dz. cyt.

[37] *Global Study on Homicide 2013*, Vienna 2014, http://www.unodc.org/documents/gsh/pdfs/2014_GLOBAL_HOMICIDE_BOOK_web.pdf .

[38] ACLU, *The war on marijuana in black and white*, czerwiec 2013, www.aclu.org/files/assets/aclu-thewaronmarijuana-ve12.pdf

[39] Może się to wydać zaskakujące, ale stanowisko firmy Equivant poparł Sąd Najwyższy stanu Wisconsin. Kiedy Eric Loomis został skazany na sześć lat pozbawienia wolności przez sędziego, który korzystał z algorytmu oceny ryzyka COMPAS, zaskarżył wyrok w wyższej instancji. W sprawie „Loomis przeciwko stanowi Wisconsin" padł zarzut, że wykorzystanie własnościowego, zamkniętego programu oceny ryzyka do przygotowania wyroku naruszyło prawa Loomisa do uczciwego procesu, ponieważ obrona nie mogła zweryfikować naukowej poprawności wyniku. Sąd Najwyższy stanu Wisconsin orzekł jednak, że wykorzystanie przez sąd w orzeczeniu algorytmicznego narzędzia do oceny ryzyka nie naruszyło praw oskarżonego do uczciwego procesu: Sąd Najwyższy stanu Wisconsin,

sprawa nr 2015AP157-CR, opinia wydana 13 lipca 2016, https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690 .

[40] Lucy Ward, *Why are there so few female maths professors in universities?*, „Guardian", 11 marca 2013, https://www.theguardian.com/lifeandstyle/the-womensblog-with-jane-martinson/2013/mar/11/women-maths-professors-uk-universities .

[41] Sonja B. Starr, M. Marit Rehavi, *Racial Disparity in Federal Criminal Charging and its Sentencing Consequences*, Ann Arbor 7 maja 2012, Program in Law and Economics Working Paper nr 12-002, http://economics.ubc.ca/files/2013/05/pdf_paper_marit-rehavi-racial-disparity-federal-criminal.pdf .

[42] David Arnold, Will Dobbie, Crystal S. Yang, *Racial Bias in Bail Decisions*, Cambridge, MA 2017, NBER Working Paper nr 23421, https://www.princeton.edu/~wdobbie/fles/racialbias.pdf .

[43] John J. Donohue III, *Capital Punishment in Connecticut, 1973–2007: A Comprehensive Evaluation from 4686 Murders to One Execution*, Stanford,CA–Cambridge, MA, październik 2011, https://law.stanford.edu/wp-content/uploads/sites/default/files/publication/259986/doc/slspublic/fulltext.pdf .

[44] Adam Benforado, *Unfair: The New Science of Criminal Injustice*, New York 2015, s. 197.

[45] Sonja B. Starr, *Estimating Gender Disparities in Federal Criminal Cases*, Ann Arbor 29 sierpnia 2012, University of Michigan Law and Economics Research Paper no. 12-018, https://ssrn.com/abstract=2144002 or http://dx.doi.org/10.2139/ssrn.2144002 .

[46] David B. Mustard, *Racial, ethnic, and gender disparities in sentencing: evidence from the US federal courts*, „Journal of Law and Economics" 2001, t. 44, nr 2, s. 285–314, http://people.terry.uga.edu/mustard/sentencing.pdf .

[47] Daniel Kahneman, *Pułapki myślenia. O myśleniu szybkim i wolnym*, tłum. Piotr Szymczak, Poznań 2012, s. 62.

[48] Chris Guthrie, Jeffrey J. Rachlinski, Andrew J. Wistrich, *Blinking on the Bench: How Judges Decide Cases*, New York 2007, paper no. 917, http://scholarship.law.cornell.edu/facpub/917 .

[49] Por. Daniel Kahneman, *Pułapki myślenia*, dz. cyt., s. 62–68.

[50] Tamże, s. 553.

[51] Mandeep K. Dhami, Peter Ayton, *Bailing and jailing the fast and frugal way*, dz. cyt.

[52] Brian Wansink, Robert J. Kent, Stephen J. Hoch, *An anchoring and adjustment model of purchase quantity decisions*, „Journal of Marketing Research" 1998, t. 35, s. 71–81, http://foodpsychology.cornell.edu/sites/default/files/unmanaged_files/Anchoring-JMR-1998.pdf .

[53] Mollie Marti, Roselle Wissler, *Be careful what you ask for: the effect of anchors on personal injury damages awards*, „Journal of Experimental Psychology: Applied" 2000, t. 6, nr 2, s. 91–103.

[54] Birte Englich, Thomas Mussweiler, *Sentencing under uncertainty: anchoring effects in the courtroom*, „Journal of Applied Social Psychology" 2001, t. 31, nr 7, s. 1535–1551, http://onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.2001.tb02687.x .

[55] Birte Englich, Thomas Mussweiler, Fritz Strack, *Playing dice with criminal sentences: the influence of irrelevant anchors on experts' judicial decision making*, „Personality and Social Psychology Bulletin" 2006, t. 32, s. 188–200, https://www.researchgate.net/publication/7389517_Playing_Dice_With_Criminal_Sentences_The_Influence_of_Irrelevant_Anchors_on_Experts%27_Judicial_Decision_Making?enrichId=rgreq-f2fedfeb71aa83f8fad80cc24df3254d-XXX&enrichSource=Y292ZXJQYWdlOzczODk1MTc7QVM6MTAzODIzNjI

wMTgyMDIyQDE0MDE3NjQ4ODgzMTA%3D&el=1_x_3&_esc=publicationCoverPdf .

[56] Tamże.

[57] Tamże.

[58] Mandeep K. Dhami, Ian K. Belton, Elizabeth Merrall, Andrew McGrath, Sheila Bird, *Sentencing in doses: is individualized justice a myth?*, w trakcie oceny. Informacje udostępnione w prywatnej korespondencji dzięki uprzejmości Mandeep Dhami.

[59] Tamże.

[60] Adam N. Glynn, Maya Sen, *Identifying judicial empathy: does having daughters cause judges to rule for women's issues?*, „American Journal of Political Science" 2015, t. 59, nr 1, s. 37–54, https://scholar.harvard.edu/files/msen/files/daughters.pdf .

[61] Shai Danziger, Jonathan Levav, Liora Avnaim-Pesso, *Extraneous factors in judicial decisions*, „Proceedings of the National Academy of Sciences of the United States of America" 2011, t. 108, nr 17, s. 6889–6892, http://www.pnas.org/content/108/17/6889 .

[62] Keren Weinshall-Margel, John Shapard, *Overlooked factors in the analysis of parole decisions*, „Proceedings of the National Academy of Sciences of the United States of America" 2011, t. 108, nr 42, E833, http://www.pnas.org/content/108/42/E833.long .

[63] Uri Simonsohn, Francesca Gino, *Daily horizons: evidence of narrow bracketing in judgment from 9,000 MBA-admission interviews*, „Psychological Science" 2013, t. 24, nr 2, s. 219–224, https://ssrn.com/abstract=2070623 .

[64] Lawrence E. Williams, John A. Bargh, *Experiencing physical warmth promotes interpersonal warmth*, „Science" 2008, t. 322, nr 5901, s. 606–607, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2737341/ .

Medycyna

[1] Richard M. Levenson, Elizabeth A. Krupinski, Victor M. Navarro, Edward A. Wasserman, *Pigeons* (Columba livia) *as trainable observers of pathology and radiology breast cancer images*, „PLOSOne", 18 listopada 2015, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0141357 .

[2] *Hippocrates' daughter as a dragon kills a knight, in* The Travels of Sir John Mandeville, British Library Online Gallery, 26 marca 2009, http://www.bl.uk/onlinegallery/onlineex/illmanus/harlmanucoll/h/011hrl000003954u00008v00.html .

[3] Eleni Tsiompanou, *Hippocrates: timeless still*, „JLL Bulletin: Commentaries on the history of treatment evaluation" 2012, http://www.jameslindlibrary.org/articles/hippocrates-timeless-still/ .

[4] David K. Osborne, *Hippocrates: father of medicine*, Greek Medicine.net, 2015, http://www.greekmedicine.net/whos_who/Hippocrates.html .

[5] Richard Colgan, *Is there room for art in evidence-based medicine?*, „Virtual Mentor", styczeń 2011, t. 13, nr 1, s. 52–54, http://journalofethics.ama-assn.org/2011/01/msoc1-1101.html .

[6] Joseph Needham, *Science and Civilization in China*, t. 6: *Biology and Biological Technology*, cz. VI: *Medicine*, red. Nathan Sivin, Cambridge 2004, s. 143, https://monoskop.org/images/1/16/Needham_Joseph_Science_and_Civilisation_in_China_Vol_6-6_Biology_and_Biological_Technology_Medicine.pdf .

[7] *Ignaz Semmelweis*, Brought to Life: Exploring the History of Medicine, http://broughttolife.sciencemuseum.org.uk/broughttolife/people/ignazsemmelweis .

[8] Wypowiedzi Andy'ego Becka pochodzą z moich prywatnych rozmów.

[9] Joann G. Elmore i in., *Diagnostic concordance among pathologists interpreting breast biopsy specimens*, „Journal of the American Medical Association", 17 marca 2015, t. 313, nr 11, s. 1122–1132, https://jamanetwork.com/journals/jama/fullarticle/2203798 .

[10] Tamże.

[11] Nazwa „sieć neuronowa" to nawiązanie do tego, co dzieje się w mózgu, gdzie miliony neuronów łączą się ze sobą w gigantyczną sieć. Każdy neuron „wsłuchuje się" w swoje połączenia i kiedy tylko wykrywa wzbudzenie innego neuronu, wysyła sygnał. Ten z kolei wzbudza kolejny neuron, który wsłuchuje się w niego.

Sztuczna sieć neuronowa to znacznie prostsza i bardziej uporządkowana wersja naszego mózgu. Tutaj (sztuczne) neurony tworzą strukturę warstwową i wszystkie neurony jednej warstwy wsłuchują się we wszystkie neurony warstwy poprzedniej. W naszym przykładzie z rozpoznawaniem psa na zdjęciu pierwsza warstwa sztucznej sieci neuronowej to pojedyncze piksele zdjęcia. Później jest wiele warstw liczących po tysiące neuronów, a na samym końcu, w ostatniej warstwie, jest tylko jeden neuron, który określa stopień prawdopodobieństwa występowania psa na badanym zdjęciu.

Procedura uaktualniania neuronów sztucznej sieci nazywana jest algorytmem propagacji wstecznej. Rozpoczynamy od ostatniego neuronu, który określa stopień prawdopodobieństwa. Powiedzmy, że załadowaliśmy do algorytmu zdjęcie psa i maszyna uznała, iż jest pewna na 70 procent, że to pies. Algorytm bada sygnały, które otrzymał od poprzedniej warstwy, i stwierdza: „Jeżeli następnym razem otrzymam takie informacje, zwiększę stopień prawdopodobieństwa występowania psa na zdjęciu". Potem maszyna informuje wszystkie neurony poprzedniej warstwy: „Ej, gdybyście dały mi taki sygnał, to moja prognoza byłaby lepsza". Każdy neuron przygląda się swoim sygnałom wejściowym i dostosowuje do tego swoje przyszłe sygnały. Potem instruuje poprzednią warstwę, jaki sygnał powinna wysłać – i tak dalej, aż do pierwszej warstwy. To właśnie temu procesowi propagacji błędów przez całą sieć od końca do początku maszyna zawdzięcza swoją nazwę „algorytmu propagacji wstecznej".

Bardziej szczegółowe omówienie sieci neuronowych, sposobu ich konstruowania i trenowania w: Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, New York 2015.

[12] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, *ImageNet classification with deep convolutional neural networks*, w: *Advances in Neural Information Processing Systems 25*, red. F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger, La Jolla, CA 2012, s. 1097–1105, http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf . Ten szczególny rodzaj algorytmu określa się mianem konwolucyjnej sieci neuronowej. Maszyna nie ogląda od razu całego zdjęcia, lecz stosuje najpierw rozmaite filtry, które wyszukują miejscowe wzorce deformacji obrazu.

[13] Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, *„Why should I trust you?" Explaining the predictions of any classifier*, w: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco 2016, s. 1135–1144, http://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf .

[14] W porównaniu z diagnozą panelu ekspertów, których zbiorcza analiza została uznana za empiryczną podstawę stanu wszystkich skrawków.

[15] Trafton Drew, Melissa L.H. Vo, Jeremy M. Wolfe, *The invisible gorilla strikes again: sustained inattentional blindness in expert observers*, „Psychological Science", wrzesień 2013, t. 24, nr 9, s. 1848–1853, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3964612/ .

[16] Goryl jest widoczny w szczycie prawego płuca.

[17] Yun Liu i in., *Detecting cancer metastases on gigapixel pathology images*, Cornell University Library, 8 marca 2017, https://arxiv.org/abs/1703.02442 .

[18] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, Andrew H. Beck, *Deep learning for identifying metastatic breast cancer*, Cornell University Library, 18 czerwca 2016, https://arxiv.org/abs/1606.05718 .

[19] David A. Snowdon, *The Nun Study*, „Boletin de LAZOS de la Asociación Alzheimer de Monterrey" 2000, t. 4, nr 22; tenże, *Healthy aging and dementia: findings from the Nun Study*, „Annals of Internal Medicine", wrzesień 2003, t. 139, nr 5, s. 450–454.

[20] „Gęstość idei" – inne określenie złożoności językowej – to suma liczby niepowtarzalnych idei przypadających na ciąg dziesięciu słów użytych przez zakonnice. Dobre omówienie tematu zob.: Associated Press, *Study of nuns links early verbal skills to Alzheimer's*, „Los Angeles Times", 21 lutego 1996, http://articles.latimes.com/1996-02-21/news/mn-38356_1_alzheimer-nuns-studied .

[21] Maja Nielsen, Jørn Jensen, Johan Andersen, *Pre-cancerous and cancerous breast lesions during lifetime and at autopsy: a study of 83 women*, „Cancer" 1984, t. 54, nr 4, s. 612–615, http://onlinelibrary.wiley.com/wol1/doi/10.1002/1097-0142(1984)54:4%3C612::AID-CNCR2820540403%3E3.0.CO;2-B/abstract .

[22] H. Gilbert Welch, William C. Black, *Using autopsy series to estimate the disease reservoir for ductal carcinoma in situ of the breast: how much more breast cancer can we find?*, „Annals of Internal Medicine", grudzień 1997, t. 127, nr 11, s. 1023–8, www.vaoutcomes.org/papers/Autopsy_Series.pdf.

[23] Uzyskanie dokładnych danych statystycznych jest trudne, bo wynik zależy od kraju i demografii (i od tego, jak intensywnie prowadzi się w danym kraju badania przesiewowe). Dobre podsumowanie zob.: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer .

[24] Cytaty pochodzą z prywatnych rozmów z Jonathanem Kanevskym.

[25] *Breakthrough method predicts risk of DCIS becoming invasive breast cancer*, „Artemis", maj 2010, http://www.hopkinsbreastcenter.org/artemis/201005/3.html .

[26] H. Gilbert Welch, Philip C. Prorok, A. James O'Malley, Barnett S. Kramer, *Breast-cancer tumor size, overdiagnosis, and mammography screening*

effectiveness, „New England Journal of Medicine" 2016, t. 375, s. 1438–1447, http://www.nejm.org/doi/full/10.1056/NEJMoa1600249 .

[27] Independent UK Panel on Breast Cancer Screening, *The benefits and harms of breast cancer screening: an independent review*, „Lancet", 30 października 2012, t. 380, nr 9855, s. 1778–1786, http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)61611-0/abstract .

[28] Prywatna korespondencja.

[29] Andrew H. Beck i in., *Systematic analysis of breast cancer morphology uncovers stromal features associated with survival*, „Science Transitional Medicine", 19 grudnia 2014, https://becklab.hms.harvard.edu/files/becklab/files/sci_transl_med-2011-beck-108ra113.pdf .

[30] Phi Vu Tran, *A fully convolutional neural network for cardiac segmentation in short-axis MRI*, 27 kwietnia 2017, https://arxiv.org/pdf/1604.00494.pdf .

[31] *Emphysema*, Imaging Analytics, Zebra Medical, https://www.zebra-med.com/algorithms/lungs/ .

[32] Eun-Jae Lee i in., *Deep into the brain: artificial intelligence in stroke imaging*, „Journal of Stroke" 2017, t. 19, nr 3, s. 277–85, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5647643/ .

[33] Taylor Kubota, *Deep learning algorithm does as well as dermatologists in identifying skin cancer*, „Stanford News", 25 stycznia 2017, https://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/ .

[34] Jo Best, *IBM Watson: the inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next*, Tech Republic, n.d., https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/ .

[35] Jennings Brown, *Why everyone is hating on IBM Watson, including the people who helped make it*, „Gizmodo", 14 sierpnia 2017, https://www.gizmodo.com.au/2017/08/why-everyone-is-hating-on-watsonincluding-the-people-who-helped-make-it/ .

[36] https://www.theregister.co.uk/2017/02/20/watson_cancerbusting_trial_on_hold_after_damning_audit_report/ .

[37] Casey Ross, Ike Swetlitz, *IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close*, „STAT", 5 września 2017, https://www.statnews.com/2017/09/05/watson-ibmcancer/ .

[38] Tomoko Otake, *Big data used for rapid diagnosis of rare leukemia case in Japan*, „Japan Times", 11 sierpnia 2016, https://www.japantimes.co.jp/news/2016/08/11/national/science-health/ibm-big-data-used-for-rapid-diagnosis-of-rare-leukemia-case-in-japan/#.Wf8S_hOoMQ8 .

[39] *Researchers validate five new genes responsible for ALS*, „Science Daily", 1 grudnia 2017, https://www.sciencedaily.com/releases/2017/12/171201104101.htm .

[40] John Freedman, *A reality check for IBM's AI ambitions*, „MIT Technology Review", 27 czerwca 2017.

[41] *Asthma facts and statistics*, Asthma UK, 2016, https://www.asthma.org.uk/about/media/facts-and-statistics/ ; *Asthma in the US*, Centers for Disease Control and Prevention, maj 2011, https://www.cdc.gov/vitalsigns/asthma/index.html .

[42] *Schoolgirl, 13, who died of asthma attack was making regular trips to A&E and running out of medication – but was NEVER referred to a specialist even when her lips turned blue, mother tells inquest*, „Daily Mail", 13 października 2015, http://www.dailymail.co.uk/news/article-3270728/Schoolgirl-13-died-asthma-attack-not-referred-specialist-lips-turned-blue.html .

[43] *My Data, My Care: How Better Use of Data Improves Health and Wellbeing*, London 2017, https://richmondgroupofcharities.org.uk/publications .

[44] Terence Carney, *Regulation 28: report to prevent future deaths, coroner's report on the case of Tamara Mills*, 29 października 2015, https://www.judiciary.gov.uk/publications/tamara-mills/ .

[45] Jamie Grierson, Alex Hern, *Doctors using Snapchat to send patient scans to each other, panel finds*, „Guardian", 5 lipca 2017, https://www.theguardian.com/technology/2017/jul/05/doctors-using-snapchat-to-send-patient-scans-to-each-other-panel-finds .

[46] Nawet jeśli jakoś sobie poradzi z tymi trudnościami, czasem po prostu nie będziemy mieli żadnych danych dla komputera. Istnieją tysiące rzadkich chorób, wyjątkowych z racji niepowtarzalnego podłoża genetycznego. Lekarze mają ogromne trudności z ich rozpoznawaniem, ponieważ nigdy wcześniej nie widzieli pacjentów, którzy by na nie chorowali. Wszystkie algorytmy świata nie rozwiążą problemu znikomości danych.

[47] Hal Hodson, *Revealed: Google AI has access to huge haul of NHS patient data*, „New Scientist", 29 kwietnia 2016, https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/ .

[48] Faktycznie najwięcej gromów z powodu tego „prawnie niewłaściwego" kontraktu spadło na fundację Royal Free Trust, która trochę zbyt gorliwie pchała się do współpracy z najsłynniejszą na świecie firmą od sztucznej inteligencji. Zob. list od Fiony Caldicott, Krajowej Strażniczki Danych, który wyciekł do Sky News: Alex Martin, *Google received 1.6 million NHS patients' data on an inappropriate legal basis*, Sky News, 15 maja 2017, https://photos.google.com/share/AF1QipMdd5VTK0RNQ1AC3Dda1526CMG0vPD4P3x4x6_qmj0Zf101rbKyxfkfyputSPvqdA/photo/AF1QipP1_rnJMXkRyy3IuFHasilQHYEknKgnHFOFEy4T?key=U2pZUDM4bmo5RHhKYVptaDlkbEhfVFh4Rm1iVUVR .

[49] Denis Campbell, *Surgeons attack plans to delay treatment to obese patients and smokers*, „Guardian", 29 listopada 2016, https://www.theguardian.com/society/2016/nov/29/surgeons-nhs-delay-treatment-obese-patients-smokers-york .

[50] Nir Eyal, *Denial of treatment to obese patients: the wrong policy on personal responsibility for health*, „International Journal of Health Policy and Management", sierpień 2013, t. 1, nr 2, s. 107–110, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3937915/ .

[51] Opis procedur badawczych można znaleźć tu: http://galton.org/essays/1880-1889/galton-1884-jaigi-anthro-lab.pdf .

[52] Francis Galton, *On the Anthropometric Laboratory at the late international health exhibition*, „Journal of the Anthropological Institute of Great Britain and Ireland" 1885, t. 14, s. 205–221.

[53] *Taste*, https://permalinks.23andme.com/pdf/samplereport_traits.pdf .

[54] *Sneezing on summer solstice?*, 23andMeBlog, 20 czerwca 2012, https://blog.23andme.com/health-traits/sneezing-on-summer-solstice/ .

[55] *Find out what your DNA says about your health, traits and ancestry*, 23andMe, https://www.23andme.com/en-gb/dna-health-ancestry/ .

[56] Kristen v. Brown, *23andMe is selling your data but not how you think*, „Gizmodo", 14 kwietnia 2017, https://gizmodo.com/23andme-is-selling-your-data-but-not-how-you-think-1794340474 .

[57] Michael Grothaus, *How 23andMe is monetizing your DNA*, Fast Company, 15 stycznia 2015, https://www.fastcompany.com/3040356/what-23andme-is-doing-with-all-that-dna .

[58] Rob Stein, *Found on the Web, with DNA: a boy's father*, „Washington Post", 13 listopada 2005, http://www.washingtonpost.com/wp-dyn/content/article/2005/11/12/AR2005111200958.html .

[59] Z analizy swojej próbki DNA mężczyzna dowiedział się, że określony wzór chromosomu Y – przekazywanego z ojca na syna – mają również dwaj mężczyźni o tym samym nazwisku (byli to jego dalecy krewni ze strony ojca). Nazwisko to, w połączeniu z miejscem i datą urodzenia ojca, wystarczyło, żeby go zidentyfikować.

[60] M. Gymrek, A.L. McGuire, D. Golan, E. Halperin, Y. Erlich, *Identifying personal genomes by surname inference*, „Science", styczeń 2013, t. 339, nr 6117, s. 321–324, https://www.ncbi.nlm.nih.gov/pubmed/23329047 .

[61] Obecnie testy genetyczne wykrywające predyspozycje do pląsawicy Huntingtona nie są dostępne w ofercie żadnego komercyjnego laboratorium.

[62] Matthew Herper, *23andMe rides again: FDA clears genetic tests to predict disease risk*, „Forbes", 6 kwietnia 2017, https://www.forbes.com/sites/matthewherper/2017/04/06/23andme-rides-again-fda-clears-genetic-tests-to-predict-disease-risk/#302aea624fdc .

Samochody

[1] DARPA, *Grand Challenge 2004: Final Report*, Arlington, VA 2004, http://www.esd.whs.mil/Portals/54/Documents/FOID/Reading_Room/DARPA/15-F-0059_GC_2004_FINAL_RPT_7-30-2004.pdf .

[2] *The Worldwide Guide to Movie Locations*, 7 września 2014, http://www.movie-locations.com/movies/k/Kill_Bill_Vol_2.html#.WkYiqrTQoQ8 .

[3] Mariella Moon, *What you need to know about DARPA, the Pentagon's mad science division*, Engadget, 7 lipca 2014, https://www.engadget.com/2014/07/07/darpa-explainer/ .

[4] DARPA, *Urban Challenge: Overview*, http://archive.darpa.mil/grandchallenge/overview.html .

[5] Sebastian Thrun, *Winning the DARPA Grand Challenge, 2 August 2006*, YouTube, 8 października 2007, https://www.youtube.com/watch?v=j8zj5lBpFTY .

[6] DARPA, *Urban Challenge: Overview*, dz. cyt.

[7] *DARPA Grand Challenge 2004 – road to...*, YouTube, 22 stycznia 2014, https://www.youtube.com/watch?v=FaBJ5sPPmcI .

[8] Alex Davies, *An oral history of the DARPA Grand Challenge, the grueling robot race that launched the self-driving car*, „Wired", 8 marca 2017, https://www.wired.com/story/darpa-grand-challenge-2004-oral-history/ .

[9] *Desert race too tough for robots*, BBC News, 15 marca 2004, http://news.bbc.co.uk/1/hi/technology/3512270.stm .

[10] Alex Davies, *An oral history of the DARPA Grand Challenge...*, dz. cyt.

[11] Denise Chow, *DARPA and drone cars: how the US military spawned self-driving car revolution*, „LiveScience", 21 marca 2014, https://www.livescience.com/44272-darpa-self-driving-car-revolution.html .

[12] Joseph Hooper, *From Darpa Grand Challenge 2004 DARPA's debacle in the desert*, „Popular Science", 4 czerwca 2004, https://www.popsci.com/scitech/article/2004-06/darpa-grand-challenge-2004darpas-debacle-desert .

[13] Alex Davies, *An oral history of the DARPA Grand Challenge...*, dz. cyt.

[14] DARPA, *Report to Congress: DARPA Prize Authority. Fiscal Year 2005 Report in Accordance with 10 U.S.C. 2374a*, marzec 2006, http://archive.darpa.mil/grandchallenge/docs/grand_challenge_2005_report_to_congress.pdf .

[15] Alan Ohnsman, *Bosch and Daimler to partner to get driverless taxis to market by early 2020s*, „Forbes", 4 kwietnia 2017,

https://www.forbes.com/sites/alanohnsman/2017/04/04/bosch-and-daimler-partner-to-get-driverless-taxis-to-market-by-early-2020s/#306ec7e63c4b .

[16] *Ford, Looking Further: Ford Will Have a Fully Autonomous Vehicle in Operation by 2021,* https://corporate.ford.com/innovation/autonomous-2021.html .

[17] John Markoff, *Should your driverless car hit a pedestrian to save your life?*, „New York Times", 23 czerwca 2016, https://www.nytimes.com/2016/06/24/technology/should-your-driverless-car-hit-a-pedestrian-to-save-your-life.html .

[18] Clive Thompson i in., *Full tilt: when 100 per cent of cars are autonomous*, „New York Times", 8 listopada 2017, https://www.nytimes.com/interactive/2017/11/08/magazine/tech-design-autonomous-future-cars-100-percent-augmented-reality-policing.html#the-end-of-roadkill .

[19] Peter Campbell, *Trucks headed for a driverless future: unions warn that millions of drivers' jobs will be disrupted*, „Financial Times", 31 stycznia 2018, https://www.ft.com/content/7686ea3e-e0dd-11e7-a0d4-0944c5f49e46 .

[20] Markus Maurer, J. Christian Gerdes, Barbara Lenz, Hermann Winner, *Autonomous Driving: Technical, Legal and Social Aspects*, New York 2016, s. 48.

[21] Stephen Zavestoski, Julian Agyeman, *Incomplete Streets: Processes, Practices, and Possibilities*, London 2015, s. 29.

[22] Markus Maurer, J. Christian Gerdes, Barbara Lenz, Hermann Winner, *Autonomous Driving*, dz. cyt., s. 53.

[23] David Rooney, *Self-guided Cars*, Science Museum, London, 27 sierpnia 2009, https://blog.sciencemuseum.org.uk/self-guided-cars/ .

[24] Blake Z. Rong, *How Mercedes sees into the future*, „Autoweek", 22 stycznia 2014, http://autoweek.com/article/car-news/how-mercedes-sees-future .

[25] Dean A. Pomerleau, *ALVINN: An Autonomous Land Vehicle In a Neural Network*, CMU-CS-89-107, Pittsburgh 1989, http://repository.cmu.edu/cgi/viewcontent.cgi?article=2874&context=compsci .

[26] Joshua Davis, *Say hello to Stanley*, „Wired", 1 stycznia 2006, https://www.wired.com/2006/01/stanley/ ; więcej szczegółów czytelnik znajdzie w: Dean A. Pomerleau, *Neural Network Perception for Mobile Robot Guidance*, New York 2012, s. 52.

[27] A. Filgueira, H. González-Jorge, S. Lagüela, L. Diaz-Vilariño, P. Arias, *Quantifying the influence of rain in LiDAR performance*, „Measurement", styczeń 2017, t. 95, s. 143–148, DOI: https://doi.org/10.1016/j.measurement.2016.10.009 ; https://www.sciencedirect.com/science/article/pii/S0263224116305577 .

[28] Chris Williams, *Stop lights, sunsets, junctions are tough work for Google's robo-cars*, „The Register", 24 sierpnia 2016, https://www.theregister.co.uk/2016/08/24/google_self_driving_car_problems /.

[29] Novatel, *IMU Errors and Their Effects*, https://www.novatel.com/assets/Documents/Bulletins/APN064.pdf .

[30] Twierdzenie Bayesa to zwykłe równanie matematyczne, które łączy stopień prawdopodobieństwa hipotezy w odniesieniu do obserwowanych informacji ze stopniem prawdopodobieństwa owych informacji w odniesieniu do hipotezy. Bardziej szczegółowe wprowadzenie do twierdzenia Bayesa można znaleźć tu: https://arbital.com/p/bayes_rule/?l=1zq .

[31] Sharon Bertsch McGrayne, *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*, New Haven 2011.

[32] M. Bayes, M. Price, *An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in*

a Letter to John Canton, A.M.F.R.S. (1763), kopia cyfrowa załadowana do archive.org 2 sierpnia 2011, https://archive.org/details/philtrans09948070 .

[33] Michael Taylor, *Self-driving Mercedes-Benzes will prioritize occupant safety over pedestrians*, „Car and Driver", 7 października 2016, https://blog.caranddriver.com/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/ .

[34] Jason Kottke, *Mercedes' solution to the trolley problem*, Kottke.org, 24 października 2016, https://kottke.org/16/10/mercedes-solution-to-the-trolley-problem .

[35] Jean-François Bonnefon, Azim Shariff, Iyad Rahwan, *The social dilemma of autonomous vehicles*, „Science", 24 czerwca 2016, t. 35, DOI: 10.1126/science.aaf2654; https://arxiv.org/pdf/1510.03346.pdf .

[36] Wszystkie cytaty z Paula Newmana pochodzą z prywatnych rozmów ze mną.

[37] Naaman Zhou, *Volvo admits its self-driving cars are confused by kangaroos*, „Guardian", 1 lipca 2017, https://www.theguardian.com/technology/2017/jul/01/volvo-admits-its-self-driving-cars-are-confused-by-kangaroos .

[38] Wszystkie wypowiedzi Jacka Stilgoego pochodzą z prywatnych rozmów ze mną.

[39] Jeff Sabatini, *The one simple reason nobody is talking realistically about driverless cars*, „Car and Driver", październik 2017, https://www.caranddriver.com/features/the-one-reason-nobody-is-talking-realistically-about-driverless-cars-feature .

[40] William Langewiesche, *The human factor*, „Vanity Fair", 17 września 2014, https://www.vanityfair.com/news/business/2014/10/air-france-flight-447-crash .

[41] Bureau d'Enquêtes et d'Analyses pour la Sécuritié de l'Aviation Civile, *Final Report on the Accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France Flight AF447 Rio de Janeiro – Paris*, wersja angielska (Paryż, uaktualniona lipiec 2012), https://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf .

[42] Tamże.

[43] William Langewiesche, *The human factor*, dz. cyt.

[44] Tamże.

[45] Jeff Wise, *What really happened aboard Air France 447*, „Popular Mechanics", 6 grudnia 2011, http://www.popularmechanics.com/flight/a3115/what-really-happened-aboard-air-france-4476611877/ .

[46] William Langewiesche, *The human factor*, dz. cyt.

[47] Jeff Wise, *What really happened aboard Air France 447*, dz. cyt.

[48] Lisanne Bainbridge, *Ironies of automation*, „Automatica", grudzień 1983, t. 19, nr 6, s. 775–779, https://www.sciencedirect.com/science/article/pii/0005109883900468 .

[49] Tamże.

[50] Alex Davies, *Everyone wants a level 5 self-driving car – here's what that means*, „Wired", 26 lipca 2016.

[51] Justin Hughes, *Car autonomy levels explained*, „The Drive", 3 listopada 2017, http://www.thedrive.com/sheetmetal/15724/what-are-these-levels-of-autonomy-anyway .

[52] Lisanne Bainbridge, *Ironies of automation*, dz. cyt.

[53] Jack Stilgoe, *Machine learning, social learning and the governance of self-driving cars*, „Social Studies of Science" 2017, t. 48, nr 1, s. 25–56.

[54] Eric Tingwall, *Where are autonomous cars right now? Four systems tested*, „Car and Driver", październik 2017, https://www.caranddriver.com/features/where-are-autonomous-cars-right-now-four-systems-tested-feature .

[55] Tracey Lindeman, *Using an orange to fool Tesla's autopilot is probably a really bad idea*, „Motherboard", 16 stycznia 2018, https://motherboard.vice.com/en_us/article/a3na9p/tesla-autosteer-orange-hack .

[56] Daisuke Wakabayashi, *Uber's self-driving cars were struggling before Arizona Crash*, „New York Times", 23 marca 2018, https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html .

[57] Sam Levin, *Video released of Uber self-driving crash that killed woman in Arizona*, „Guardian", 22 marca 2018, https://www.theguardian.com/technology/2018/mar/22/video-released-of-uber-self-driving-crash-that-killed-woman-in-arizona .

[58] Audi, *The Audi vision of autonomous driving*, Audi Newsroom, 11 września 2017, https://media.audiusa.com/en-us/releases/184 .

[59] P. Morgan, C. Alford, G. Parkhurst, *Handover Issues in Autonomous Driving: A Literature Review. Project Report*, Bristol 2016, http://eprints.uwe.ac.uk/29167/1/Venturer_WP5.2Lit_ReviewHandover.pdf .

[60] William Langewiesche, *The human factor*, dz. cyt.

[61] Evan Ackerman, *Toyota's Gill Pratt on self-driving cars and the reality of full autonomy*, „IEEE Spectrum", 23 stycznia 2017, https://spectrum.ieee.org/cars-that-think/transportation/self-driving/toyota-gill-pratt-on-the-reality-of-full-autonomy .

[62] Julia Pyper, *Self-driving cars could cut greenhouse gas pollution*, „Scientific American", 15 września 2014,

https://www.scientificamerican.com/article/self-driving-cars-could-cut-greenhouse-gas-pollution/ .

[63] Raphael E. Stern i in., *Dissipation of stop-and-go waves via control of autonomous vehicles: field experiments*, arXiv: 1705.01693v1, 4 maja 2017, https://arxiv.org/abs/1705.01693 .

[64] SomeJoe7777, *Tesla Model S forward collision warning saves the day*, YouTube, 19 października 2016, https://www.youtube.com/watch?v=SnRp56XjV_M .

[65] Jordan Golson, Dieter Bohn, *All new Tesla cars now have hardware for full self-driving capabilities: but some safety features will be disabled initially*, „The Verge", 19 października 2016, https://www.theverge.com/2016/10/19/13340938/tesla-autopilot-update-model-3-elon-musk-update .

[66] Fred Lambert, *Tesla introduces first phase of Enhanced Autopilot: measured and cautious for next several hundred million miles" – release notes*, Electrek, 1 stycznia 2017, https://electrek.co/2017/01/01/tesla-enhanced-autopilot-release-notes/ .

[67] DPC Cars, *Toyota Guardian and Chauffeur autonomous vehicle platform*, YouTube, 27 września 2017, https://www.youtube.com/watch?v=IMdceKGJ9Oc .

[68] Brian Milligan, *The most significant development since the safety belt*, BBC News, 15 kwietnia 2018, http://www.bbc.co.uk/news/business-43752226 .

Przestępczość

[1] Bob Taylor, *Crimebuster: Inside the Minds of Britain's Most Evil Criminals*, London 2002, rozdz. 9: *A day out from jail*.

[2] Tamże.

[3] Nick Davies, *Dangerous, in prison – but free to rape*, „Guardian", 5 października 1999, https://www.theguardian.com/uk/1999/oct/05/nickdavies1 .

[4] João Medeiros, *How geographic profiling helps find serial criminals*, „Wired", listopad 2014, http://www.wired.co.uk/article/mapping-murder .

[5] *The Origins of Criminology: A Reader*, red. Nicole H. Rafter, Abingdon 2009, s. 271.

[6] Luke Dormehl, *The Formula: How Algorithms Solve All Our Problems... and Create More*, London 2014, s. 117.

[7] Tamże, s. 116.

[8] D. Kim Rossmo, *Geographic profiling*, w: *Encyclopedia of Criminology and Criminal Justice*, red. Gerben Bruinsma, David Weisburd, New York 2014, https://link.springer.com/referenceworkentry/10.1007%2F978-1-4614-5690-2_678 .

[9] Tamże.

[10] João Medeiros, *How geographic profiling helps find serial criminals*, dz. cyt.

[11] Tamże.

[12] *Sadistic serial rapist sentenced to eight life terms*, „Independent" (Ireland), 6 października 1999, http://www.independent.ie/worldnews/sadistic-serial-rapist-sentenced-to-eight-life-terms26134260.html .

[13] Tamże.

[14] Steven C. Le Comber, D. Kim Rossmo, Ali N. Hassan, Douglas O. Fuller, John C. Beier, *Geographic profiling as a novel spatial tool for targeting infectious disease control*, „International Journal of Health Geographics" 2011, t. 10, nr 1, s. 35, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3123167/ .

[15] Michelle V. Hauge, Mark D. Stevenson, D. Kim Rossmo, Steven C. Le Comber, *Tagging Banksy: using geographic profiling to investigate a modern art mystery*, „Journal of Spatial Science" 2016, t. 61, nr 1, s. 185–90, http://www.tandfonline.com/doi/abs/10.1080/14498596.2016.1138246 .

[16] Raymond Dussault, *Jack Maple: betting on intelligence*, „Government Technology", 31 marca 1999, http://www.govtech.com/featured/Jack-Maple-Betting-on-Intelligence.html .

[17] Tamże.

[18] Tamże.

[19] Nicole Gelinas, *How Bratton's NYPD saved the subway system*, „New York Post", 6 sierpnia 2016, http://nypost.com/2016/08/06/how-brattons-nypd-saved-the-subway-system/ .

[20] Raymond Dussault, *Jack Maple: betting on intelligence*, dz. cyt.

[21] Andrew Guthrie Ferguson, *Predictive policing and reasonable suspicion*, „Emory Law Journal" 2012, t. 62, nr 2, s. 259, http://law.emory.edu/elj/content/volume-62/issue-2/articles/predicting-policing-and-reasonable-suspicion.html .

[22] Lawrence W. Sherman, Patrick R. Gartin, Michael E. Buerger, *Hot spots of predatory crime: routine activities and the criminology of place*, „Criminology" 1989, t. 27, nr 1, s. 27–56, http://onlinelibrary.wiley.com/doi/10.1111/j.1745-9125.1989.tb00862.x/abstract .

[23] Toby Davies, Shane D. Johnson, *Examining the relationship between road structure and burglary risk via quantitative network analysis*, „Journal of Quantitative Criminology" 2015, t. 31, nr 3, s. 481–507, http://discovery.ucl.ac.uk/1456293/5/Johnson_art%253A10.1007%252Fs10940-014-9235-4.pdf .

[24] Michael J. Frith, Shane D. Johnson, Hannah M. Fry, *Role of the street network in burglars' spatial decision-making*, „Criminology" 2017, t. 55, nr 2,

s. 344–76, http://onlinelibrary.wiley.com/doi/10.1111/1745-9125.12133/full .

[25] Spencer Chainey, *JDI Brief: Predictive Mapping (Predictive Policing)*, London 2012, http://discovery.ucl.ac.uk/1344080/3/JDIBriefs_PredictiveMappingSChaineyApril2012.pdf .

[26] Tamże.

[27] Mały *disclaimer*: algorytm PredPolu nie jest publicznie dostępny. Eksperyment, o którym tu opowiadamy, przeprowadzili ci sami matematycy, którzy założyli PredPol, używając techniki pasującej do opisu własnościowego oprogramowania algorytmu. Wszystko wskazuje na to, że to jedno i to samo, nie mamy jednak absolutnej pewności.

[28] G.O. Mohler i in., *Randomized controlled field trials of predictive policing*, „Journal of the American Statistical Association" 2015, t. 110, nr 512, s. 1399–1411, http://www.tandfonline.com/doi/abs/10.1080/01621459.2015.1077710 .

[29] Kent Police Corporate Services Analysis Department, *PredPol Operational Review*, 2014, http://www.statewatch.org/docbin/uk-2014-kent-police-predpol-op-review.pdf .

[30] G.O. Mohler i in., *Randomized controlled field trials of predictive policing*, dz. cyt.

[31] Kent Police Corporate Services Analysis Department, *PredPol Operational Review: Initial Findings*, 2013, https://www.whatdotheyknow.com/request/181341/response/454199/attach/3/13_10_888_Appendix.pdf .

[32] Tamże.

[33] Nie użyto wówczas PredPola, lecz znacznie prostszy algorytm, który również wykorzystywał zjawiska sygnału i wzmocnienia. Zob. Matthew Fielding, Vincent Jones, *Disrupting the optimal forager: predictive risk mapping*

and domestic burglary reduction in Trafford, Greater Manchester, „International Journal of Police Science and Management" 2012, t. 14, nr 1, s. 30–41.

[34] Joe Newbold, *Predictive policing, preventative policing or intelligence led policing. What is the future?*, projekt konsultacyjny przedstawiony do oceny w programie Warwick MBA, Warwick Business School, 2015.

[35] Dane z 2016: COMPSTAT, *Citywide Profile 12/04/16–12/31/16*, http://assets.lapdonline.org/assets/pdf/123116cityprof.pdf .

[36] Ronald V. Clarke, Mike Hough, *Crime and Police Effectiveness*, London 1984, Home Office Research Study nr 79, https://archive.org/stream/op1276605-1001/op1276605-1001_djvu.txt , relacja w: Tom Gash, *Criminal: The Truth about Why People Do Bad Things*, London 2016.

[37] Kent Police Corporate Services Analysis Department, *PredPol Operational Review*, dz. cyt.

[38] PredPol, *Recent examples of crime reduction*, 2017, http://www.predpol.com/results/ .

[39] Aaron Shapiro, *Reform predictive policing*, „Nature", 25 stycznia 2017, t. 541, nr 7638, http://www.nature.com/news/reform-predictive-policing-1.21338 .

[40] Chicago Data Portal, Strategic Subject List, https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np .

[41] Jessica Saunders, Priscilla Hunt, John Hollywood, *Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot*, „Journal of Experimental Criminology" 2016, t. 12, nr 3, s. 347–371.

[42] Copblock, *Innocent man arrested for robbery and assault, spends two months in Denver jail*, 28 kwietnia 2015, https://www.copblock.org/122644/man-

arrested-for-robbery-assault-he-didnt-commit-spends-two-months-in-denver-jail/ .

[43] Tamże.

[44] Ava Kofman, *How a facial recognition mismatch can ruin your life*, „The Intercept", 13 października 2016.

[45] Tamże.

[46] Copblock, *Denver police, „Don't f*ck with the biggest gang in Denver". Before beating man wrongfully arrested – TWICE!*, 30 stycznia 2016, https://www.copblock.org/152823/denver-police-fck-up-again/ .

[47] W rzeczywistości historia Talleya była jeszcze gorsza niż w przedstawionym w książce krótkim streszczeniu. Po spędzeniu dwóch miesięcy w więzieniu po pierwszym zatrzymaniu został wypuszczony na wolność po oddaleniu zarzutów. Rok później został aresztowany po raz drugi. Tym razem zarzutów nie oddalono i agent FBI, który obsługiwał program do rozpoznawania twarzy, zeznawał przeciw niemu. Sprawa Talleya zaczęła się ostatecznie rozpadać dopiero wtedy, kiedy kasjer z obrabowanego banku zdał sobie sprawę, że Talley nie ma brodawek, które wcześniej widział na rękach napastnika, i zeznał przed sądem: „To nie jest gość, który mnie obrabował". Obecnie Talley domaga się 10 milionów odszkodowania. Pełny opis sprawy czytelnik znajdzie w: Ava Kofman, *How a facial recognition mismatch can ruin your life*, dz. cyt.

[48] Justin Huggler, *Facial recognition software to catch terrorists being tested at Berlin station*, „Telegraph", 2 sierpnia 2017, http://www.telegraph.co.uk/news/2017/08/02/facial-recognition-software-catch-terrorists-tested-berlin-station/.

[49] David Kravets, *Driver's license facial recognition tech leads to 4,000 New York arrests*, „Ars Technica", 22 sierpnia 2017, https://arstechnica.com/tech-policy/2017/08/biometrics-leads-to-thousands-of-a-ny-arrests-for-fraud-identity-theft/ .

[50] Ruth Mosalski, *The first arrest using facial recognition software has been made*, „Wales Online", 2 czerwca 2017, http://www.walesonline.co.uk/news/local-news/first-arrest-using-facial-recognition-13126934.

[51] Sebastian Anthony, *UK police arrest man via automatic face recognition tech*, „Ars Technica", 6 czerwca 2017, https://arstechnica.com/tech-policy/2017/06/police-automatic-face-recognition .

[52] David White, Richard I. Kemp, Rob Jenkins, Michael Matheson, A. Mike Burton, *Passport officers' errors in face matching*, „PLOSOne", 18 sierpnia 2014, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0103510#s6 .

[53] Teghan Lucas, Maciej Henneberg, *Are human faces unique? A metric approach to finding single individuals without duplicates in large samples*, „Forensic Science International", grudzień 2015, t. 257, s. 514.e1–514.e6, http://www.sciencedirect.com/science/article/pii/S0379073815003758 .

[54] Zaria Gorvett, *You are surprisingly likely to have a living doppelganger*, BBC Future, 13 lipca 2016, http://www.bbc.com/future/story/20160712-you-are-surprisingly-likely-to-have-a-living-doppelganger .

[55] *Eyewitness misidentification*, The Innocence Project, https://www.innocenceproject.org/causes/eyewitness-misidentification .

[56] Douglas Starr, *Forensics gone wrong: when DNA snares the innocent*, „Science", 7 marca 2016, http://www.sciencemag.org/news/2016/03/forensics-gone-wrong-when-dna-snares-innocent .

[57] Co nie znaczy bynajmniej, że przy badaniach genetycznych fałszywe identyfikacje nie są wykluczone, bo rzeczywiście się zdarzają; oznacza to tylko tyle, że masz do dyspozycji narzędzie, które pozwala ci ograniczyć takie przypadki do absolutnego minimum.

[58] Richard W. Vorder Bruegge, *Individualization of People from Images*, Quantico, Va., FBI Operational Technology Division, Forensic Audio, Video and Image Analysis Unit, 12 grudnia 2016, https://www.nist.gov/sites/default/files/documents/2016/12/12/vorderbruegge-face.pdf .

[59] Lance Ulanoff, *The iPhone X can't tell the difference between twins*, Mashable UK, 31 października 2017, http://mashable.com/2017/10/31/putting-iphone-x-face-id-to-twin-test/#A87kA26aAqqQ .

[60] Kif Leswing, *Apple says the iPhone X's facial recognition system isn't for kids*, „Business Insider UK", 27 września 2017, http://uk.businessinsider.com/apple-says-the-iphone-xs-face-id-is-less-accurate-on-kids-under-13-2017-9 .

[61] Andy Greenberg, *Watch a 10-year-old's face unlock his mom's iPhone X*, „Wired", 14 listopada 2017, https://www.wired.com/story/10-year-old-face-id-unlocks-mothers-iphone-x/ .

[62] *Bkav's new mask beats Face ID in „twin way": severity level raised, do not use Face ID in business transactions*, Bkav Corporation, 27 listopada 2017, http://www.bkav.com/dt/top-news/-/view_content/content/103968/bkav%EF%BF%BDs-new-mask-beats-face-id-in-twin-way-severity-level-raised-do-not-use-face-id-in-business-transactions .

[63] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, Michael Reiter, *Accessorize to a crime: real and stealthy attacks on state-of-the-art face recognition*, artykuł zaprezentowany na konferencji ACM SIGSAC Conference, 2016, https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf .

[64] Ira Kemelmacher-Shlizerman, Steven M. Seitz, Daniel Miller, Evan Brossard, *The MegaFace Benchmark: 1 Million Faces for Recognition at Scale*, Computer Vision Foundation, 2015, https://arxiv.org/abs/1512.00596 .

[65] *Half of all American adults are in a police face recognition database, new report finds*, informacja prasowa, Georgetown Law, 18 października 2016, https://www.law.georgetown.edu/news/press-releases/half-of-all-american-adults-are-in-a-police-face-recognition-database-new-report-finds.cfm .

[66] Josh Chin, Liza Lin, *China's all-seeing surveillance state is reading its citizens' faces*, „Wall Street Journal", 6 czerwca 2017, https://www.wsj.com/articles/the-all-seeing-surveillance-state-feared-in-the-west-is-a-reality-in-china-1498493020 .

[67] Daniel Miller, Evan Brossard, Steven M. Seitz, Ira Kemelmacher-Shlizerman, *The MegaFace Benchmark: 1 Million Faces for Recognition at Scale*, 2015, https://arxiv.org/pdf/1505.02108.pdf .

[68] Tamże.

[69] *MegaFace and MF2: Million-Scale Face Recognition: most recent public results*, 12 marca 2017, http://megaface.cs.washington.edu/ ; *Leading facial recognition platform Tencent YouTu Lab smashes records in MegaFace facial recognition challenge*, Cision PR Newswire, 14 kwietnia 2017, http://www.prnewswire.com/news-releases/leading-facial-recognition-platform-tencent-youtu-lab-smashes-records-in-megaface-facial-recognition-challenge-300439812.html .

[70] Dan Robson, *Facial recognition a system problem gamblers can't beat?*, TheStar.com, 12 stycznia 2011, https://www.thestar.com/news/gta/2011/01/12/facial_recognition_a_system_problem_gamblers_cant_beat.html .

[71] British Retail Consortium, *2016 Retail Crime Survey: February 2017*, London 2017, https://brc.org.uk/media/116348/10081-brc-retail-crime-survey-2016_all-graphics-latest.pdf .

[72] *The D&D Daily's 2016 Retail Violent Death Report*, D&D Daily, 9 marca 2017, http://www.d-ddaily.com/archivesdaily/DailySpecialReport03-09-17F.htm .

[73] Joan Gurney, *Walmart's use of facial recognition tech to spot shoplifters raises privacy concerns*, iQ Metrix, 9 listopada 2015, http://www.iqmetrix.com/blog/walmarts-use-of-facial-recognition-tech-to-spot-shoplifters-raises-privacy-concerns .

[74] Andy Coghlan, James Randerson, *How far should fingerprints be trusted?*, „New Scientist", 14 września 2005, https://www.newscientist.com/article/dn8011-how-far-should-fingerprints-be-trusted/ .

[75] Phil Locke, *Blood spatter – evidence?*, The Wrongful Convictions Blog, 30 kwietnia 2012, https://wrongfulconvictionsblog.org/2012/04/30/blood-spatter-evidence/ .

[76] Michael Shermer, *Can we trust crime forensics?*, „Scientific American", 1 września 2015, https://www.scientificamerican.com/article/can-we-trust-crime-forensics/ .

[77] National Research Council of the National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward*, Washington, DC 2009, s. 7, https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf .

[78] Colin Moynihan, *Hammer attacker sentenced to 22 years in prison*, „New York Times", 19 lipca 2017, https://www.nytimes.com/2017/07/19/nyregion/hammer-attacker-sentenced-to-22-years-in-prison.html?mcubz=0 .

[79] Jeremy Tanner, *David Baril charged in hammer attacks after police-involved shooting*, Pix11, 14 maja 2015, http://pix11.com/2015/05/14/david-baril-charged-in-hammer-attacks-after-police-involved-shooting/ .

[80] *Long-time fugitive captured juggler was on the run for 14 years*, FBI, 12 sierpnia 2014, https://www.fbi.gov/news/stories/long-time-fugitive-neil-stammer-captured .

[81] Pei-Sze Cheng, *I-Team: use of facial recognition technology expands as some question whether rules are keeping up*, NBC 4NewYork, 23 czerwca 2015, http://www.nbcnewyork.com/news/local/Facial-Recognition-NYPD-Technology-Video-Camera-Police-Arrest-Surveillance-309359581.html .

[82] Nate Berg, *Predicting crime, LAPD-style*, „Guardian", 25 czerwca 2014, https://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report .

Sztuka

[1] Matthew J. Salganik, Peter Sheridan Dodds, Duncan J. Watts, *Experimental study of inequality and unpredictability in an artificial cultural market*, „Science", 10 lutego 2006, t. 311, s. 854, DOI: 10.1126/science.1121066, https://www.princeton.edu/~mjs3/salganik_dodds_watts06_full.pdf .

[2] http://www.princeton.edu/~mjs3/musiclab.shtml .

[3] Kurt Kleiner, *Your taste in music is shaped by the crowd*, „New Scientist", 9 lutego 2006, https://www.newscientist.com/article/dn8702-your-taste-in-music-is-shaped-by-the-crowd/ .

[4] Bjorn Carey, *The science of hit songs*, LiveScience, 9 lutego 2006, https://www.livescience.com/7016-science-hit-songs.html .

[5] *Vanilla, indeed*, True Music Facts Wednesday Blogspot, 23 lipca 2014, http://truemusicfactswednesday.blogspot.co.uk/2014/07/tmfw-46-vanilla-indeed.html .

[6] Matthew J. Salganik, Duncan J. Watts, *Leading the herd astray: an experimental study of self-fulfilling prophecies in an artificial cultural market*, „Social Psychology Quarterly", jesień 2008, t. 74, nr 4, s. 338, DOI: https://doi.org/10.1177/019027250807100404 .

[7] S. Sinha, S. Raghavendra, *Hollywood blockbusters and long-tailed distributions: an empirical study of the popularity of movies*, „European Physical

Journal B" 2004, t. 42, s. 293–6, DOI: https://doi.org/10.1140/epjb/e2004-00382-7 ; http://econwpa.repec.org/eps/io/papers/0406/0406008.pdf .

[8] *John Carter: analysis of a so-called flop: a look at the box office and critical reaction to Disney's early tentpole release John Carter,* WhatCulture, http://whatculture.com/film/john-carter-analysis-of-a-so-called-flop .

[9] J. Valenti, *Motion pictures and their impact on society in the year 2000,* przemówienie wygłoszone w Midwest Research Institute, Kansas City, 25 kwietnia 1978, s. 7.

[10] William Goldman, *Adventures in the Screen Trade,* New York 1983.

[11] Sameet Sreenivasan, *Quantitative analysis of the evolution of novelty in cinema through crowdsourced keywords,* Scientific Reports 3, artykuł nr 2758, 2013, uaktualniony 29 stycznia 2014, DOI: https://doi.org/10.1038/srep02758 , https://www.nature.com/articles/srep02758 .

[12] Márton Mestyán, Taha Yasseri, János Kertész, *Early prediction of movie box office success based on Wikipedia activity big data,* „PLOSOne", 21 sierpnia 2013, DOI: https://doi.org/10.1371/journal.pone.0071226 .

[13] Ramesh Sharda, Dursun Delen, *Predicting box-office success of motion pictures with neural networks,* „Expert Systems with Applications" 2006, t. 30, nr 2, s. 243–4, DOI: https://doi.org/10.1016/j.eswa.2005.07.018 ; https://www.sciencedirect.com/science/article/pii/S0957417405001399 .

[14] Banksy NY, *Banksy sells work for $60 in Central Park, New York – video,* „Guardian", 14 października 2013, https://www.theguardian.com/artanddesign/video/2013/oct/14/banksy-central-park-new-york-video .

[15] Bonhams, *Lot 12 Banksy: Kids on Guns,* 2 lipca 2014, http://www.bonhams.com/auctions/21829/lot/12/ .

[16] Charlie Brooker, *Supposing… subversive genius Banksy is actually rubbish*, „Guardian", 22 września 2006, https://www.theguardian.com/commentisfree/2006/sep/22/arts.visualarts .

[17] Gene Weingarten, *Pearls before breakfast: can one of the nation's greatest musicians cut through the fog of a DC rush hour? Let's find out*, „Washington Post", 8 kwietnia 2007, https://www.washingtonpost.com/lifestyle/magazine/pearls-before-breakfast-can-one-of-the-nations-great-musicians-cut-through-the-fog-of-a-dc-rush-hour-lets-find-out/2014/09/23/8a6d46da-4331-11e4-b47c-f5889e061e5f_story.html?utm_term=.a8c9b9922208 .

[18] Cytaty z Armanda Leroi pochodzą z prywatnej korespondencji. Badania, o których mówi Leroi, to: Matthias Mauch, Robert M. MacCallum, Mark Levy, Armand M. Leroi, *The evolution of popular music: USA 1960–2010*, „Royal Society Open Science", 6 maja 2015, DOI: https://doi.org/10.1098/rsos.150081 .

[19] Cytaty z Davida Cope'a pochodzą z prywatnej korespondencji.

[20] Cytat został skrócony. Zob. Douglas Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, London 1979, s. 673.

[21] George Johnson, *Undiscovered Bach? No, a computer wrote it*, „New York Times", 11 listopada 1997.

[22] *Autobiography of Mark Twain*, red. Benjamin Griffin, Harriet Elinor Smith, t. 3, Oakland, CA–London 2015), cz. 1, s. 103.

[23] Lew Tołstoj, *Co to jest sztuka?*, tłum. Maria Leśniewska, Kraków 1980, s. 91.

[24] Douglas Hofstadter, *Gödel, Escher, Bach*, dz. cyt., s. 674.

Zakończenie

[1] Historię Rahinah Ibrahim opisano tu: https://www.propublica.org/article/fbi-checked-wrong-box-rahinah-ibrahim-terrorism-watch-list ; https://alumni.stanford.edu/get/page/magazine/article/?article_id=66231 .

[2] GenPact, *Don't underestimate importance of process in coming world of AI*, 14 lutego 2018, http://www.genpact.com/insight/blog/dont-underestimate-importance-of-process-in-coming-world-of-ai .

[1*] Przypisy oznaczone cyframi znajdują się na s. 261–297.

[2*] A wszystko przez to, że nazwa miejscowości zawiera ciąg liter „cunt", uważany za bardzo nieparlamentarne określenie żeńskich narządów płciowych (przyp. tłum.).

[3*] To parafraza słów, jakimi obawy przed SI skomentował informatyk i pionier uczenia maszynowego Andrew Ng w dyskusji z 2015 r. (Tech Events, *GPU Technology Conference 2015 day 3: What's Next in Deep Learning*, YouTube, 20 listopada 2015) https://www.youtube.com/watch?v=qP9TOX8T-kI .

[4*] Stworzenie symulacji działania mózgu nicienia jest celem międzynarodowego projektu naukowego o nazwie OpenWorm. Jego uczestnicy mają nadzieję odtworzyć sieć 302 neuronów mózgu nicienia *Caenorhabditis elegans*. Dla porównania – my, ludzie, mamy około 100 000 000 000 neuronów. Zob. stronę internetową OpenWorm: http://openworm.org/ .

[5*] Intrygujący i rzadki wyjątek od tej reguły wyższości algorytmów nad człowiekiem został poświadczony przez wiele przeprowadzonych na przełomie lat 50. i 60. badań „diagnozowania" (sformułowanie badaczy, nie moje) homoseksualizmu. Przewidywania człowieka w tym zakresie biły na głowę najlepsze dokonania algorytmów – co mogłoby wskazywać na istnienie spraw immanentnie ludzkich, z którymi operujące na twardych danych formuły matematyczne zawsze będą miały ogromny problem.

[6*] Reklamy to nie jedyny powód ich istnienia. Ciasteczek używają również strony internetowe do kontrolowania, czy jesteś zalogowany (żeby wiedzieć, czy można ci bezpiecznie przesyłać informacje wrażliwe) albo czy to twoje pierwsze, czy już kolejne odwiedziny na stronie (żeby na przykład podnieść cenę biletu na samolot lub przesłać ci e-mailem kod promocyjny do internetowego sklepu z odzieżą).

[7*] Jak na ironię, w warunkach użytkowania wtyczki – przewrotnie nazwanej The Web of Trust (Sieć zaufania) – wszystkie te informacje były

zapisane czarno na białym.

[8*] Taka kombinacja sugeruje, że publikowałabym więcej tweetów, gdybym nie przejmowała się tak bardzo tym, jak zostaną odebrane przez innych.

[9*] Podobny wynik możesz uzyskać nawet wówczas, kiedy nie zakodujesz płci do swojego algorytmu. Jeżeli tylko prognoza opiera się na czynnikach, które bardziej korelują z jedną, a nie z drugą grupą (takich jak kryminalna przeszłość oskarżonego), jest możliwe, że maszyna potraktuje je nierówno.

[10*] Gdyby piłka kosztowała 10 centów, to kij musiałby kosztować 1 dolara i 10 centów, a to już daje razem 1 dolara i 20 centów.

[11*] W Polsce odpowiednikiem był *Va banque!* (przyp. tłum.).

[12*] Więcej o Bayesie w rozdziale *Samochody*.

[13*] Faktycznie nie da się ustalić, czy jesteś wikingiem, czy nie – co próbował wyjaśnić mi zawile mój dobry znajomy, genetyk Adam Rutherford. Wspominam o tym głównie po to, żeby go sprowokować. Aby zrozumieć, dlaczego nie jest to możliwe z punktu widzenia nauki, trzeba przeczytać jego książkę: *Krótka historia wszystkich ludzi, którzy kiedykolwiek żyli. Opowieści zapisane w naszych genach* (tłum. Adam Tuz, Warszawa 2017).

[14*] Watson, maszyna diagnostyczna skonstruowana przez firmę IBM, o której mówiliśmy w rozdziale *Medycyna*, w dużym stopniu wykorzystuje wnioskowanie bayesowskie. Zob. https://www.ibm.com/developerworks/library/os-ind-watson/ .

[15*] Zwycięzcę wyścigu w 2005 roku, pojazd zespołu Uniwersytetu Stanforda, krótko i zwięźle opisał stanfordzki matematyk Persi Diaconis: „To auto było bayesowskie po najmniejszą śrubkę".

[16*] W prasie można było przeczytać kilka różnych wersji tego samego scenariusza: „A gdyby przechodniem była 90-letnia staruszka? Albo małe

dziecko? A gdyby w samochodzie jechał zdobywca Nagrody Nobla?". Wszystkie sprowadzają się jednak do tego samego dylematu.

[17*] Problemom związanym z brakiem praktyki można oczywiście zaradzić. Od czasów katastrofy samolotu Air France w szkoleniu zawodowym dużo uwagi poświęca się umiejętności sterowania po awarii autopilota, a już latający piloci regularnie wyłączają autopilota, żeby nie stracić wprawy.

[18*] Asystent do jazdy w korkach to kolejny krok w drodze do automatyzacji – poziom 3. Może przejmować kierowanie samochodem w ściśle określonych sytuacjach. Człowiek nadal musi być gotowy do interwencji, w razie gdyby samochód napotkał scenariusz, którego nie rozumie, nie musi jednak bezustannie czuwać, obserwując drogę i inne samochody. Ten poziom bardziej przypomina nakłanianie nastolatka do umycia naczyń.

[19*] Kiedy piszę tę książkę, w lutym 2018 roku, owo „pełne oprzyrządowanie potrzebne do samosterowania" to płatna opcja dodatkowa, którą można nabyć podczas zakupu samochodu, choć tesle nie mają obecnie oprogramowania niezbędnego do jeżdżenia w trybie pełnej autonomiczności. Na stronie Tesli czytamy: „Nie jesteśmy w stanie dokładnie powiedzieć, kiedy będzie dostępny każdy z elementów opisanej wyżej funkcjonalności". Zob. https://www.tesla.com/en_GB/blog/all-tesla-cars-being-produced-now-have-full-self-driving-hardware .

[20*] Mowa o autentycznym produkcie o nazwie The Autopilot Buddy, który można kupić za rewelacyjną cenę 179 dolarów. Warto dodać, że na stronie internetowej napisano małym drukiem: „Autopilot BuddyTM nie powinien być nigdy używany na drogach publicznych". https://www.autopilotbuddy.com/ .

[21*] W tym przypadku piszę „wiemy" jak najbardziej dosłownie – mam bowiem na myśli także samą siebie; akurat przy tym badaniu współpracowałam z moim wybitnym doktorantem Michaelem Frithem.

[22*] Liczba ta pokazuje, ile par zdjęć tej samej twarzy algorytm nie zdołał ze sobą powiązać, kiedy został ustawiony na unikanie błędnych rozpoznań. Więcej o rodzajach błędów popełnianych przez algorytmy można przeczytać w rozdziale *Wymiar sprawiedliwości*, a o różnych sposobach mierzenia ich dokładności – w rozdziale *Medycyna*.

[23*] Z drugiej strony, w wieku 14 lat byłam wielką fanką chłopców z PJ&Duncan, więc trudno powiedzieć…

[24*] Bardzo żałuję, że format tej książki nie pozwala na załączenie fragmentów muzyki, bo tego po prostu t r z e b a posłuchać – piosenka jest tak beznadziejna, że aż chce się śmiać. Koniecznie znajdźcie ją w sieci!

[25*] Polecam też gorąco zapoznanie się z muzyką Cope'a w sieci. Moim ulubionym kawałkiem jest chyba utwór na orkiestrę w stylu Vivaldiego: https://www.youtube.com/watch?v=2kuY3BrmTfQ .

[26*] Możesz również wykorzystać tę funkcję telefonu do napisania tekstu imitującego twój styl – otwórz nową notatkę i wpisz parę słów, żeby uruchomić algorytm, na przykład „Urodziłam się…". A potem już tylko stukaj w dowolne słowa z listy sugestii wyświetlanych na ekranie. Oto moja (autentyczna) próbka. Zaczyna się nieźle, ale pod koniec robi się trochę nerwowo: „(Urodziłam się) żeby być dobrą osobą i chciałabym być z tobą bardzo wiele ludzi wiem że nie dostajesz moich maili i nie mam na to czasu".

[27*] Cope zauważył również, że w miarę postępu pracy algorytmu należało pilnować kilku innych miar, takich jak długość frazy i długość utworu, które okazały się kluczem do udanej imitacji Bacha.

[28*] Umówiłam się z dyrektorem generalnym tej spółki na rozmowę i zapytałam go, czy kiedykolwiek sprawdził, czy jego algorytmy rzeczywiście to potrafią. W odpowiedzi usłyszałam długą anegdotę o tym, jak analiza dokonana przez jego sieć neuronową doprowadziła do zwolnienia wielkiej gwiazdy Hollywoodu z udziału w pewnej serii filmów. Kiedy zauważyłam, że to raczej dowód na to, iż ludzie kupują jego

marketing, a nie na to, że algorytm działa, odpowiedział: „Badania akademickie to nie nasza branża”.

[29*] Jest jedna sztuczka pozwalająca wyłapać takie szemrane algorytmy. Nazywam ją „testem magii”. Następnym razem, kiedy zobaczysz reklamę jakiegoś algorytmu, sprawdź, co się stanie, jeżeli zamienisz dowolny slogan – „uczenie maszynowe”, „sztuczna inteligencja” czy „sieć neuronowa” – słowem „magia” albo „czary”. Czy zdania nadal trzymają się kupy? Czy zachowała się istotna część przekazu? Jeżeli nie, to zdaje się, że coś tu bardzo śmierdzi. Obawiam się, że nawet w bardzo odległej przyszłości nie zdołamy „rozwiązać problemu głodu na świecie za pomocą czarów” ani „stosować magii do pisania idealnych scenariuszy”.